Polygenic score application to complex trait prediction

Selzam, Saskia

Awarding institution:
King's College London

Polygenic Score Application to Complex Trait Prediction

Saskia Selzam

Social, Genetic and Developmental Psychiatry Centre, King's College London, University of London

Submitted for the degree of Doctor of Philosophy in Statistical Genetics

2019

Acknowledgements

My PhD was funded in part by a UK Medical Research Council (MRC) studentship to the Social, Genetic and Developmental Psychiatry Centre at the Institute of Psychiatry, Psychology and Neuroscience (IoPPN), King's College London, by the IoPPN itself, and by the US National Institutes of Health (AG046938). The research uses data collected as part of the Twins Early Development Study (TEDS), supported by programme grant to Professor Robert Plomin from the UK Medical Research Council (MR/M021475/1 and previously G0901245), with additional support from the US National Institutes of Health (AG046938). The research leading to the results in this thesis has also received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/ grant agreement n° 602768 and ERC grant agreement n° 295366. High performance computing facilities used to process and analyse data in relation to this research were funded with capital equipment grants from the GSTT Charity (TR130505) and Maudsley Charity (980).

I would like to express my deep gratitude to my supervisor Professor Robert Plomin for his support, encouragement, guidance and generosity throughout my PhD. I am also very grateful to my supervisors Dr. Paul O'Reilly and Dr. Jonathan Coleman, who have been instrumental in my development of methodological and statistical skills.

It has been an absolute privilege to be part of the Twins Early Development Study team, and I am immensely thankful to Rachel Ogden, Andrew McMillan and Louise Webster for their help and support. I feel fortunate for having been surrounded by so many talented and inspiring individuals throughout this PhD journey, many of whom have become close friends - thank you for having made my every day more fun and enjoyable.

Finally, I am very grateful to my parents and their partners for their trust and support in my choices. Thanks to my siblings for the joy and laughter they bring. To my partner James – I cannot thank you enough for always being there, in your kind and loving way.

Author Declaration

Data used for this research were collected as part of the Twins Early Development Study (TEDS), which is an ongoing longitudinal study that recruited twin pairs from England and Wales between 1994 – 1996. Phenotypic data in Chapter 2, 3, 5 and part of Chapter 6 were collected prior to the research described here. I contributed to the collection of some measures in Chapter 6. Individual-level genotypic data in Chapter 2 was processed and quality controlled by Dr. Eva Krapohl, a previous TEDS PhD student. For Chapter 3, 4, 5, and 6, I quality controlled, imputed and harmonized individual-level genotypic data. For Chapters 2, 3, 5, and 6, I was responsible for preparing all phenotypic and genetic measures. For Chapter 4, I prepared polygenic score data in TEDS, and used secondary data based on previous publications. I was responsible for all analyses presented in this thesis. To the best of my knowledge, the work presented here is original and my own work, except where acknowledged in the text.

Saskia Selzam

Abstract

Decades of quantitative genetics research has led to the conclusion that all human behavioural traits tested to date show genetic influence to varying degrees. However, powerful individual-level genetic prediction based on measured genetic variation has only become feasible more recently through technological and methodological advancements. Especially over the past decade, the application of genetic prediction methods has grown exponentially, permeating research in the social, behavioural and biomedical sciences.

This thesis seeks to investigate developmental and multivariate research questions, as well as genotype-environment interplay through the use of a broad range of polygenic scores analysis approaches and phenotypes. The research sample in this work was the UK Twins Early Development Study (TEDS), utilised to investigate the predictive accuracy of a genome-wide polygenic score (GPS) for years of education for a variety of cognitive and non-cognitive traits (Chapter 2 and 3); the shared genetic aetiology between a range of psychiatric disorders (Chapter 4); potential genotype-environment interactions (GxE) for education and intelligence using the polygenic score for years of education (Chapter 2); the influence of evocative genotype-environment correlation (rGE) in the relationship between child body-mass index and parental feeding practices (Chapter 5); and finally, the effect of passive rGE on polygenic score prediction estimates across various trait domains by comparing within- and between-family polygenic score predictions (Chapter 6).

This thesis provided evidence that 1) target trait prediction estimates of the GPS for years of education represent the strongest polygenic prediction of any behavioural trait; 2) the GPS for years of education is associated with a wide range of traits, including cognition, personality, BMI, physical and mental health, and environmental measures; 3) the substantial genetic overlap between psychiatric disorders may be due to a common genetic factor; 4) there is no evidence for GxE for cognitive and educational outcomes using the GPS for years of education; and 5) evocative and passive rGE can be detected and quantified through GPS analysis.

Leveraging a broad range of phenotypes and GPS, this thesis illustrates the usefulness of the polygenic score approach in research. General implications, limitations and future directions will be discussed (Chapter 7).

Index

Index of Figures

Index of Tables

Chapter 3 – Polygenic score for educational attainment captures DNA variants shared between personality traits and educational achievement

Chapter 1 – General Introduction

Decades of quantitative genetics research has led to the conclusion that all human behavioural traits tested to date show genetic influence to varying degrees (Knopik, Neiderhiser, DeFries, & Plomin, 2017; Polderman et al., 2015; Turkheimer, 2000), however, powerful individual-level genetic prediction based on measured genetic variation has only become tangible more recently through technological and methodological advancements. Especially over the past decade, the application of genetic prediction methods has grown exponentially, permeating research in the social, behavioural and biomedical sciences. This progress has opened up novel opportunities to investigate developmental and multivariate research questions, as well as genotype-environment interplay, and it is these investigations that represent the broad key aims of this thesis.

It is the aim of Chapter 1 to provide an overview of human complex trait genetics and genome-wide association analyses to offer context to the polygenic scoring method, with an emphasis on recent advancements in the field.

The polygenic model for quantitative traits

In 1866, Mendel discovered distinct inheritance patterns for discrete pea plant traits, which accurately predicted categorical trait outcome probabilities in offspring generations (Mendel, 1866). His principles of genetic inheritance became highly influential in the understanding of genetic transmission of monogenic traits in humans, where the trait outcome is entirely dependent on the inheritance of dominant or recessive alleles for a single genetic marker. However, monogenic traits make up only a small fraction of all human traits, and Mendel's laws appeared too simplistic at first to explain the generation of quantitative traits (Lander & Kruglyak, 1995; Nadeau, 2001). Quantitative traits are characterised by their variation in a population, for which individuals can display any value on a continuous scale.

In 1918, the biometrician Fisher proposed the now widely applied infinitesimal model, stating that all genetic variants with small effect each contribute to the development of quantitative traits (Fisher, 1918), as opposed to single genetic markers as previously suggested. This model assumes that although genetic transmission is based on Mendelian inheritance patterns, genetic effects operate additively rather than through dominant or recessive mechanisms. This polygenic model of inheritance suggests that when aggregating

multiple trait-associated alleles, the combined effect allele frequency distribution will approach a normal distribution based on the central limit theorem (Falconer & MacKay, 1996). This expectation translates to few individuals carrying either a small or a large number of trait increasing alleles, and to most individuals carrying close to the average number of trait increasing alleles in a population. Quantitative traits are also often referred to as complex traits, as quantitative trait variation is typically the result of highly polygenic architecture and multifaceted environmental influences (Neale, Ferreira, Medland, & Posthuma, 2008).

Before genome-wide genotyping technologies became available, family studies, including twin and adoption studies, were the exclusive approaches for the estimation of overall genetic and environmental influences in trait variability (Knopik et al., 2017). Using expectations about genetic relatedness between family members based on biometrical genetics theory, it is possible to derive a measure of genetic influence through contrasting trait resemblance between family members of varying degrees of genetic relatedness (Mather & Jinks, 1977). Genetic influence can be described by the variance component *narrow-sense heritability* (h^2), which is the proportion of phenotypic variability that is caused by additive genetic variability in a population:

$$h^2 = \frac{\sigma_A^2}{\sigma_A^2 + \sigma_D^2 + \sigma_C^2 + \sigma_E^2}$$

Equation 1. Definition of heritability.

with A, D, C and E being additive and dominant genetic effects, shared environmental and non-shared environmental effects, respectively. The denominator represents the total phenotypic variance.

Thousands of quantitative genetics studies conducted over the past 50 years have led to the overall conclusion that all complex human traits studied to date show genetic influence (Knopik et al., 2017; Polderman et al., 2015; Turkheimer, 2000). Although there has been recent progress, it has proven difficult in statistical genetics research to reproduce $h^2_{pedigree}$ estimates from family studies based on directly measured genetic variation – also known as the *missing heritability* problem (Manolio et al., 2009).

Genetic association and prediction

One of the first attempts to identify genomic regions related to specific trait outcomes were linkage studies, which typically study pedigrees that show a pattern of disease penetrance (Neale et al., 2008). The success of linkage analysis is limited to true Mendelian disorders where effect sizes of culprit genes are large (Glazier, Nadeau, & Aitman, 2002). Due to the polygenic architecture of complex traits, where effects of individual genetic variants are additive and small, linkage analysis is severely underpowered to detect any of these effects (Risch & Merikangas, 1996; Teare & Barrett, 2005).

With the advent of technological advances, the focus of genetic research shifted to a genome-wide approach. The International HapMap project (The International HapMap Consortium, 2003) identified the correlation patterns between genetic markers across the human genome (i.e. linkage disequilibrium; LD), consequently pinpointing genetic markers that are independent from one another. Owing to this success, modern genotyping microarrays are designed to leverage this information, obtaining maximum coverage with a relatively small number of genetic variants (Daly, Rioux, Schaffner, Hudson, & Lander, 2001; Gabriel et al., 2002; Johnson et al., 2001; The International HapMap Consortium, 2003). Although there are several types of genetic variation, such as copy number variation, insertions and deletions, microarrays predominantly capture common genetic variation (minor allele frequency (MAF) typically > 1%) using single nucleotide polymorphisms (SNPs). SNPs are the most frequent form of genetic variation in humans (The 1000 Genomes Project Consortium et al., 2015), and are single base pair changes across the genome that correspond with phenotypic differences between individuals. Due to the high LD across the human genome, statistical techniques can be employed to impute genotypes from microarray assays (Marchini & Howie, 2010) using reference panels such as from the 1000 Genomes Project (The 1000 Genomes Project Consortium et al., 2015) or the Haplotype Reference Consortium (McCarthy et al., 2016) to obtain a large coverage of SNPs for further analysis. The advantages of microarray genotyping are cost-effectiveness, efficiency, and accuracy (Syvänen, 2005). However, genotyping platforms are unable to capture rare or de novo mutations, and due to the limited selection of SNPs most likely do not include variants that are truly causal, but variants that are tagged by LD to the causal variant (Visscher et al., 2017; Wray et al., 2013).

Genome-wide association

The aim of genome-wide association (GWA) studies is to discover single SNPs that are associated with an outcome trait using a systematic approach that is agnostic to the causal variants or genes (often termed 'hypothesis-free'). SNP discoveries from GWA analysis can be informative in two main ways: they may be the foundation for further aetiological research into biological mechanisms, or may be utilised as biomarkers that tag heritable trait variance. For the purpose of this thesis, the focus will be on the latter.

For SNP discovery, genetic data can be modelled based on the assumption of dominant, recessive, or multiplicative genetic effects, but GWA studies most commonly apply an additive model. This model assumes that an increasing number of trait associated alleles relates to a linear increase in the phenotypic trait values in a population. Associations are tested between each SNP across the genome and the phenotypic trait values in the sample population, using a linear regression model for continuous traits, and a logistic model for binary outcomes (Bush & Moore, 2012). The two main broad findings from GWA analyses so far are extreme polygenicity and pleiotropy (i.e. one SNP affecting several traits) across the genome (Visscher et al., 2017; Visscher, Brown, McCarthy, & Yang, 2012a).

The influence of a SNP is evaluated based on two metrics: statistical significance and the effect size of the association. The statistical significance threshold in GWA studies is $5e^{-8}$, based on a Bonferroni correction on the number of independent SNPs across the genome to account for an inflation of Type I errors due to extensive multiple testing (Gratten, Wray, Keller, & Visscher, 2014; Jannot, Ehret, & Perneger, 2015). Genome-wide significant SNP effect sizes are much smaller than initially anticipated, typically explaining less than 0.1% of the phenotypic variance each (Visscher et al., 2012a; Visscher, Goddard, Derks, & Wray, 2012b). This general finding from GWA research highlights the extreme scale of polygenicity to the contribution of heritability.

The highly conservative significance threshold applied in GWA together with the small individual SNP effect sizes poses a challenge for obtaining adequate statistical power. Power in GWA studies depends on multiple factors, including study sample size, phenotype heterogeneity and measurement accuracy, the overall number of causal SNPs, and the distribution of these SNP effect sizes (Korte & Farlow, 2013; Visscher et al., 2017). The scale of sample sizes required has become particularly apparent through the recent accessibility to

enormous study samples. These have become possible due the availability of cost-effective and efficient genotyping technology, giving rise to large population biobanks, as well as international collaborations between academia and industry. For example, GWA meta-analyses on height led to 180 SNP discoveries for height based on 183,727 individuals, 697 SNP associations in a sample of 254,288 individuals, and 1185 height-associated SNPs in 693,529 individuals (Allen et al., 2010; Wood et al., 2014; Yengo et al., 2018). In psychiatric genetics, an increase in power due to larger GWA study samples has also led to an increase in genetic discoveries, albeit the overall number of discoveries remains relatively low. For example, a GWA study on schizophrenia including 9,394 cases and 12,462 controls yielded seven significant SNPs, a later study with 36,989 cases and 113,075 controls identified 128 SNPs, whereas the latest meta-analysis of 40,675 cases and 64,643 controls identified 179 independent SNPs (Pardiñas et al., 2018; Ripke et al., 2011; Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014). A recent GWA meta-analysis on major depressive disorder only discovered 44 SNPs, despite being the largest psychiatric genetics study with a sample of 135,458 cases and 344,901 controls (Wray et al., 2018). These findings indicate that even the currently available large sample sizes are not yet sufficient to identify all causal variants. This is likely due to population disorder prevalence and phenotype heterogeneity for some disorders (Gratten et al., 2014). Another explanation might be extremely small SNP effects sizes, and that genotyping arrays do not capture structural and rare mutations, which have been shown to influence the development of psychiatric traits (Smoller et al., 2019).

The most successful GWA study for any behavioural trait deployed a proxy-phenotype approach, which used the trait *years of education* as a marker of educational success (Rietveld et al., 2014). This approach made it possible to obtain very large sample sizes, because years of education is assessed in most GWA studies as a standard demographic variable. The aggregation and meta-analysis of this data was achieved by standardising multi-national cohort data to the International Standard Classification of Education scale (UNESCO Institute for Statistics, 2006). Since the publication on the first GWA study on years of education on 126,559 individuals in 2014 (Rietveld et al., 2014), two follow-up meta-analyses were published, including 294,723 and 1,131,881 individuals respectively (Lee et al., 2018; Okbay et al., 2016), with the latter being the largest GWA meta-analysis conducted to date for any trait. This study identified 1,271 independent SNPs, on average explaining ~0.02% of the phenotypic variance each.

Not only have GWA studies been useful in identifying and replicating SNP discoveries and in revealing the extreme level of polygenicity, they have also highlighted the vast scale of pleiotropy across human complex traits (Bulik-Sullivan, Finucane, et al., 2015b; Cross-Disorder Group of the Psychiatric Genomics Consortium, 2013; Shi, Kichaev, & Pasaniuc, 2016; Sivakumaran et al., 2011; Visscher & Yang, 2016). Pleiotropy describes the scenario where one genetic marker affects several traits. For example, an early study by the Cross Disorder Group of the Psychiatric Genetics Consortium based on a total of 33,332 cases and 27,888 controls showed four SNPs that were significantly associated with a range of psychiatric disorders, and that overall genetic risk associated with specific disorders predicted cross-disorder liability (Cross-Disorder Group of the Psychiatric Genomics Consortium, 2013). A recent cross-disorder meta-analysis of 232,964 cases and 494,162 controls of eight psychiatric traits indicated that at least two disorders significantly shared 109 pleiotropic genetic variants, and at least four disorders had 23 markers in common (Cross-Disorder Group of the Psychiatric Genomics Consortium et al., 2019).

Beyond genome-wide significant pleiotropic effects, the vast scale of pleiotropy across the genome is now widely accepted, with potentially thousands of genetic variants with small effect each contributing to the phenotypic association between traits (Bulik-Sullivan, Finucane, et al., 2015b; Docherty, Moscati, & Fanous, 2016; Shi, Mancuso, Spendlove, & Pasaniuc, 2017). The extent of genetic pleiotropy between two traits can be expressed through genetic correlation, which is the ratio of the genetic covariance and the genetic variances of the two traits:

$$r_g = \frac{cov_g}{\sqrt{\sigma^2_{A_x}\sigma^2_{A_y}}}$$

Equation 2. Definition of genetic correlation.

Where cov_g represents the genetic covariance between trait x and trait y, and $\sigma^2_{A_x}$ and $\sigma^2_{A_y}$ the additive genetic variance of trait x and trait y, respectively. If r_g= 0, there are no shared genetic effects between the traits, and if r_g= 1, this indicates a total overlap between the genetic factors. Nevertheless, it is often not possible to distinguish between biological (or horizontal) pleiotropy, where one DNA marker directly affects several traits, and mediated

(or vertical) pleiotropy, where one DNA marker directly affects one trait, which then in turn affects another trait (Solovieff, Cotsapas, Lee, Purcell, & Smoller, 2013).

Polygenic score prediction

Initially, the main goal of GWA analysis was to identify single SNPs that can be used to trace mechanisms from genes to brain to trait (eg. behaviour). Given the challenges posed by small effect sizes due to extreme polygenicity, a new approach emerged by changing the focus from statistical significance to genetic prediction in independent samples. Predictive power can be greatly increased by aggregating genetic effects estimated through GWA analysis across SNPs even if they do not individually reach genome-wide statistical significance (Dudbridge, 2013; Palla & Dudbridge, 2015). This approach is consistent with the polygenic model of inheritance (Fisher, 1918) and was first successfully applied to predict schizophrenia risk using summary data from a GWA study on schizophrenia (Purcell et al., 2009). Aggregate scores of a large number of genetic effects have been shown to be more predictive of trait outcomes than genetic instruments that include a few genome-wide significant SNPs (Okbay et al., 2016; Purcell et al., 2009; Yengo et al., 2018).

A genome-wide polygenic score (GPS) is the sum of the weighted count of the number of trait-associated alleles across the genome and calculated as:

$$GPS_i = \sum_{j=1}^{k} \hat{\beta}_j G_j$$

Equation 3. Definition of a polygenic score.

where GPS_i represents the GPS of the i_{th} individual in the target sample, which should be independent from and unrelated to the GWA (i.e. discovery) study sample, and $j \in \{1,2,..,k\}$ and indexes the k number of SNPs included in the GWA analysis. G_j reflects the genotype dosage (0, 1, or 2 alleles) of SNP j. $\hat{\beta}_j$ represents an estimate of the true effects size of SNP j as obtained through GWA analysis. This effect is measured imperfectly for various reasons, such as low statistical power for GWA analysis leading to sampling variance on genetic effect sizes, the inclusion of SNPs in the analysis that are not causal but in LD with the causal variants, or measurement error of the trait in question (Barendse, 2011; Dudbridge, 2013). Overall, the more genetic markers that are included in the GPS, the greater will be the

measurement error in the aggregate score (Dudbridge, 2016), highlighting the need for precise effect size estimates.

The statistical power of GPS prediction depends on several factors. These are mainly related to the discovery (i.e. GWA) study sample size, the genetic covariance between discovery and target sample, the heritable trait variance of the phenotype, the number of causal markers that contribute to the genetic variance, disease prevalence for binary disorders, and lastly, target sample size (Dudbridge, 2013; Palla & Dudbridge, 2015; Wray, Goddard, & Visscher, 2007). Importantly, predictive performance of the GPS relies strongly on the discovery sample size. For example, a GPS for years of education based on a discovery sample of 293,723 predicting 3.2% of the variance in educational achievement requires only 250 individuals in the target sample to achieve 80% power (Okbay et al., 2016), https://tinyurl.com/y56ec8jh).

When building a GPS, a critical step is to account for LD among the genetic variants in the target sample, as it can lead to an upward bias of the explained trait variance (Bulik-Sullivan, Loh, et al., 2015a). The effect of each genetic marker on the trait is estimated individually in GWA analysis, therefore SNPs that are in LD tag the effects of their surrounding variants. As a result, a first step to polygenic scoring is to create a thinned set of independent genetic markers. One approach is to perform *clumping*, which starts with the selection of the most significant SNP based on GWA analysis, and only retains subsequent SNPs in order of their *p*-value if their location does not fall within a specified window of previously selected SNPs. A common software for polygenic score creation that uses this approach is *PRSice* (Euesden, Lewis, & O'Reilly, 2015), which selects SNPs for inclusion in the GPS based on their *p*-value threshold. Here, the selection of *p*-value thresholds can be defined manually. Alternatively, a 'best-fit' approach can be applied, whereby many GPS are calculated between the *p*-value thresholds of 0 and 1 for any given stepwise *p*-value increment. Linear regressions are then run for each GPS and the outcome, and the 'best-fit' GPS with the highest variance explained is selected for further analysis.

A different method incorporates a Bayesian approach, implemented in the software *LDpred* (Vilhjalmsson et al., 2015). Rather than removing markers through clumping, LDpred retains all SNPs but adjusts $\hat{\beta}_j$ to account for local LD, and a prior on the effect size of each SNP given the assumed genetic architecture of the trait. This prior depends on the genetic variance of the discovery trait and an assumption about the fraction of causal markers

believed to influence the discovery trait. For example, a causal fraction of 1 implies the assumption that all SNPs causally influence the discovery trait. Therefore, the prior re-weights the $\hat{\beta}_j$ such that the effects are spread out amongst the SNPs across the whole genome in proportion to the LD present amongst these SNPs.

Applications of genome-wide polygenic scores

Target- and cross-trait prediction

One of the main applications for GPS in this thesis is association testing by relating the polygenic score variance to phenotypic variance in a sample, which was performed in all research included in Chapters 2 - 6. The strength of the association is evaluated through the coefficient of determination R^2:

$$R^2 = \frac{h^2_{SNP}}{1 + \frac{M}{Nh^2_{SNP}}(1 - R^2)}$$

Equation 4. Definition of the proportion of phenotypic variance explained by a quantitative genetic predictor.

Where R^2 represents the proportion of phenotypic variance explained by the M independent SNPs that account for the h^2_{SNP} proportion of the total phenotypic variance in a GWA study with N number of individuals (Daetwyler, Villanueva, & Woolliams, 2008; Wray et al., 2013). Equation 4 applies to target-trait prediction, where a GPS for a specific trait predicts the same phenotypic outcome, and also to cross-trait prediction, where the GPS for a specific trait predicts a different phenotypic trait. The latter approach can be used if the target GWA analysis is underpowered and a more powered GWA study is available for a correlated trait (Chapter 2 & 3), or if the aim is to investigate shared aetiology between different traits due to pleiotropy (Chapter 3 & 4) (Choi, Mak, & O'Reilly, 2018).

The predictive efficacy of the genetic predictor is assessed through:

$$R^2_{G,\hat{G}} = \frac{R^2}{h^2_{SNP}}$$

Equation 5. The GPS prediction relative to the total genetic variance.

which evaluates the phenotypic variance explained by the genetic predictor relative to the maximum genetic prediction possible as defined through h^2_{SNP}. Therefore, h^2_{SNP} represents the ceiling of GPS predictions, and is also commonly referred to as chip-h^2 because it is estimated using SNPs genotyped on microarray chips. Additive genetic variance based on SNPs can be estimated through individual-level genotypes by applying a mixed-effects linear model as implemented in GCTA-REML (Yang, Lee, Goddard, & Visscher, 2011) or LDAK (Speed et al., 2017). GWA summary statistic-based methods to estimate h^2_{SNP}, such as LDSC, have also gained popularity. This method is based on the principle that the presence of LD across the genome is correlated with the upward bias of the individual SNP test statistics as estimated through GWA analysis, and the magnitude of this correlation is indicative of polygenic signal (Bulik-Sullivan, Loh, et al., 2015a). h^2_{SNP} is generally lower than $h^2_{pedigree}$ because it only captures additive genetic effects of the common variants measured on the microarray or that are imputed, and does not include effects of rare variants or non-additive effects (epistasis, dominance). Genetic effects are currently estimated with a certain degree of error due to insufficient power in GWA analysis, which is partly why R^2 is generally smaller than h^2_{SNP}.

Investigating genotype-environment interplay using polygenic scores

The principles described above can also be applied to investigate genotype-environment interplay through polygenic score analysis. One type of genotype-environment interplay is genotype-environment correlation (rGE) (Knopik et al., 2017; Plomin, DeFries, & Loehlin, 1977), which is the concept that an individual's genetically influenced behaviour may elicit specific reactions from others (evocative rGE), or lead individuals to choose experiences and environments that correlate with their genotype (active rGE). A third type of rGE is passive rGE, whereby children are exposed to family environments that are partly created by, and therefore correlated with their parents' genetic propensities. There is converging evidence from twin and adoption studies that supports rGE as a developmental mechanism. For example, moderate heritability estimates have been shown for exposure to life events, home environments including parental warmth and discipline, television viewing, and also behaviours that influence health outcomes, such as smoking, alcohol intake, drug consumption, and other risky behaviours (Ball & Collier, 2002; Bolinskey, Neale, Jacobson,

Prescott, & Kendler, 2004; Dick & Foroud, 2003; Jaffee & Price, 2007; Li, 2003; Plomin & Bergeman, 1991; Plomin, Corley, DeFries, & Fulker, 2017; Prescott, Madden, & Stallings, 2006; Price & Jaffee, 2008; Saudino, Pedersen, Lichtenstein, McClearn, & Plomin, 1997).

However, twin studies are unable to study the genetic influence in exposures that are invariant to members of the same family, such as family socio-economic status (SES). Although adoption studies can assess genetic influence on between-family environmental measures, adoption data is often not easily accessible. Through the advent of DNA based methods in unrelated individuals, widespread evidence for rGE mechanisms emerged. For example, h^2_{SNP} estimates of family SES range between ~10 – 20%, and there is substantial genetic overlap with other traits, such as intelligence, educational achievement, physical and mental health, and anthropometric traits (Hill et al., 2016; Krapohl & Plomin, 2016; Marioni et al., 2014; Trzaskowski et al., 2014). Perhaps unsurprisingly, GPS analysis has shown that a polygenic score for years of education in the offspring predicts 2.5% of the variability in family SES and 1% of the variance in chaos at home (Krapohl et al., 2016). Another study established that the education GPS accounted for 23% of the covariance between breast feeding duration and the child's educational achievement, and 6% and 7% of the covariation between corporal punishment and children's conduct problems and attention-deficit hyperactivity problems, respectively (Krapohl et al., 2017). In Chapters 5 and 6, polygenic score analysis was applied to investigate rGE mechanisms in developmental outcomes.

The other type of genotype-environment interplay is genotype-environment interaction (GxE), which refers to the idea that the effect of the environment on the trait outcome is dependent on an individual's genotype (Knopik et al., 2017; Plomin et al., 1977; Rutter, Moffitt, & Caspi, 2006). For example, some twin studies have suggested that $h^2_{pedigree}$ estimates of childhood intelligence vary as a function of family SES, showing that $h^2_{pedigree}$ increases with family SES (Rowe, Jacobson, & Van den Oord, 1999; Tucker-Drob, Rhemtulla, Harden, Turkheimer, & Fask, 2010; Turkheimer, Haley, Waldron, D'Onofrio, & Gottesman, 2003). The interpretation is that in high SES families, genetic effects can be realized through the provision of environmental opportunities, whereas in low SES families the lack of environmental opportunities hinders the genetic potential. However, other studies have not been able to replicate this effect (Grant et al., 2010; Hanscombe et al., 2012; van der Sluis, Willemsen, de Geus, Boomsma, & Posthuma, 2008). A meta-analysis of 14 independent twin studies provided context to the inconsistency amongst findings, showing that a GxE effect

for intelligence and educational achievement was only observed in samples from the United States, possibly due to a higher gradient in social and educational inequality (Tucker-Drob & Bates, 2016). The only known GPS study that tested for a GxE effect used a polygenic score based on the first GWA for years of education (Rietveld et al., 2014), created in a US sample, and found no significant effect (Conley et al., 2015). In Chapter 2, this GxE hypothesis was tested through polygenic score analysis in a UK sample.

Chapter overview

This thesis presents work that mainly relies on the polygenic score method to 1) establish prediction accuracy for a variety of developmental outcomes, 2) to gain a better understanding of shared aetiology between traits, and 3) to test genotype-environment interplay.

With the surge in power to estimate genetic effect sizes more accurately in GWA analysis, polygenic score predictors have become increasingly stronger in their trait predictions. Chapter 2 investigated the predictive power of a GPS based on the most powerful GWA meta-analysis of years of education at the time, including ~329,000 individuals, which was a follow-up analysis of a previous study including ~126,000 participants. The outcome traits included educational achievement measured at ages 7, 12 and 16, intelligence at age 12, and family SES. This study also tested the GxE hypothesis that the effect of the education GPS on educational achievement and intelligence depends on family SES.

As an extension to the traits investigated in Chapter 2, Chapter 3 explored the predictive strength of a GPS based on a further updated GWA meta-analysis of years of education (~766,000 individuals due to exclusion of 23andme data) in relation to six personality domains that are associated with educational achievement. Prediction estimates using the education GPS were contrasted to two GPS for personality: neuroticism and wellbeing. Using structural equation modelling, this study also tested to what extent the three GPS accounted for the phenotypic covariance between the personality domains and educational achievement.

It has been proposed that a single dimension, the psychopathology or *p* factor, can capture an individual's liability to psychiatric disorders. Recent genetic research has found high genetic correlations between pairs of mental disorders. The aim in Chapter 4 was to

investigate shared genetic aetiology between psychiatric traits. The hypothesis of a genetic *p* factor was tested by applying principal component analysis to genetic correlation matrices of various psychiatric disorders, estimated through four genetic methods: polygenic scoring, LDSC, GCTA-REML, and family study.

In Chapter 5, it was the aim to explore evocative rGE mechanisms in the home environment. It is widely believed that parents influence their child's body-mass index (BMI) through certain feeding practices. Given that BMI has a strong genetic basis, an alternative hypothesis was tested: child BMI evokes parental feeding practices. A GPS for BMI was created in the offspring and used to predict measured parental restriction over food intake, and parental pressure to increase food intake. Between- and within-family analyses were employed to test associations. In addition, multivariate twin analysis was used to estimate the genetic correlations between child BMI and parental feeding practices.

It has recently been shown that a substantial proportion of the SNP effect sizes estimated in a GWA meta-analysis of years of education is attributable to passive rGE (i.e. shared family environment). As GPSs are calculated using GWA SNP effect sizes that potentially include environmental effects, Chapter 6 turns to the estimation of passive rGE effects in polygenic score prediction in a sample of genotyped DZ twin pairs. A mixed-effects model was used to account for the clustering, including two fixed effects to estimate within- and between-family simultaneously. Through this model, within- and between-family polygenic score predictions of eight core life outcomes (anthropometric, cognitive, personality and health) for eight corresponding polygenic scores were compared for both target- and cross-trait associations.

Chapter 7 concludes with an overall discussion of the limitations, implications and future direction of polygenic score analysis.

References

Allen, H. L., Estrada, K., Lettre, G., Berndt, S. I., Weedon, M. N., Rivadeneira, F., et al. (2010). Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature*, *467*(7317), 832–838. http://doi.org/10.1038/nature09410

Barendse, W. (2011). The effect of measurement error of phenotypes on genome wide association studies. *Bmc Genomics*, *12*(1). http://doi.org/10.1186/1471-2164-12-232

Bulik-Sullivan, B. K., Loh, P.-R., Finucane, H. K., Ripke, S., Yang, J., Patterson, N., et al. (2015a). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics*, *47*(3), 291–295. http://doi.org/10.1038/ng.3211

Bulik-Sullivan, B., Finucane, H. K., Anttila, V., Gusev, A., Day, F. R., Loh, P.-R., et al. (2015b). An atlas of genetic correlations across human diseases and traits. *Nature Genetics*, *47*(11), 1236–1241. http://doi.org/10.1038/ng.3406

Bush, W. S., & Moore, J. H. (2012). Genome-Wide Association Studies. *PLoS Computational Biology*, *8*(12). http://doi.org/10.1371/journal.pcbi.1002822

Choi, S. W., Mak, T. S. H., & O'Reilly, P. (2018). A guide to performing Polygenic Risk Score analyses. *bioRxiv*, 416545. http://doi.org/10.1101/416545

Conley, D., Domingue, B. W., Cesarini, D., Dawes, C., Rietveld, C. A., & Boardman, J. D. (2015). Is the Effect of Parental Education on Offspring Biased or Moderated by Genotype? *Sociological Science*, *2*(6), 82–105. http://doi.org/10.15195/v2.a6

Cross-Disorder Group of the Psychiatric Genomics Consortium. (2013). Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *The Lancet*, *381*(9875), 1371–1379. http://doi.org/10.1016/S0140-6736(12)62129-1

Cross-Disorder Group of the Psychiatric Genomics Consortium, Lee, P. H., Anttila, V., Won, H., Feng, Y.-C. A., Rosenthal, J., et al. (2019). Genome wide meta-analysis identifies genomic relationships, novel loci, and pleiotropic mechanisms across eight psychiatric disorders. *bioRxiv*, 528117. http://doi.org/10.1101/528117

Daetwyler, H. D., Villanueva, B., & Woolliams, J. A. (2008). Accuracy of Predicting the Genetic Risk of Disease Using a Genome-Wide Approach. *Plos One*, *3*(10), e3395. http://doi.org/10.1371/journal.pone.0003395

Daly, M. J., Rioux, J. D., Schaffner, S. F., Hudson, T. J., & Lander, E. S. (2001). High-resolution haplotype structure in the human genome. *Nature Genetics*, *29*(2), 229–232. http://doi.org/10.1038/ng1001-229

Docherty, A. R., Moscati, A. A., & Fanous, A. H. (2016). Cross-Disorder Psychiatric Genomics. *Current Behavioral Neuroscience Reports*, *3*(3), 256–263. http://doi.org/10.1007/s40473-016-0084-3

Dudbridge, F. (2013). Power and Predictive Accuracy of Polygenic Risk Scores. *PLOS Genetics*, *9*(3). http://doi.org/10.1371/journal.pgen.1003348

Dudbridge, F. (2016). Polygenic Epidemiology. *Genetic Epidemiology*, *40*(4), 268–272. http://doi.org/10.1002/gepi.21966

Euesden, J., Lewis, C. M., & O'Reilly, P. F. (2015). PRSice: Polygenic Risk Score software. *Bioinformatics*, *31*(9), 1466–1468. http://doi.org/10.1093/bioinformatics/btu848

Falconer, D., & MacKay, T. (1996). Introduction to Quantitative Genetics (4 ed.). Harlow, UK: Longman.

Fisher, R. A. (1918). The Correlation between Relatives on the Supposition of Mendelian Inheritance. *Transactions of the Royal Society of Edinburgh*, *52*, 399–433. http://doi.org/10.1017/S0080456800012163

Gabriel, S. B., Schaffner, S. F., Nguyen, H., Moore, J. M., Roy, J., Blumenstiel, B., et al. (2002). The structure of haplotype blocks in the human genome. *Science*, *296*(5576), 2225–2229. http://doi.org/10.1126/science.1069424

Glazier, A. M., Nadeau, J. H., & Aitman, T. J. (2002). Finding genes that underlie complex traits. *Science*, *298*(5602), 2345–2349. http://doi.org/10.1126/science.1076641

Grant, M. D., Kremen, W. S., Jacobson, K. C., Franz, C., Xian, H., Eisen, S. A., et al. (2010). Does parental education have a moderating effect on the genetic and environmental influences of general cognitive ability in early adulthood? *Behavior Genetics*, *40*(4), 438–446. http://doi.org/10.1007/s10519-010-9351-3

Gratten, J., Wray, N. R., Keller, M. C., & Visscher, P. M. (2014). Large-scale genomics unveils the genetic architecture of psychiatric disorders. *Nature Neuroscience*, *17*(6), 782–790. http://doi.org/10.1038/nn.3708

Hanscombe, K. B., Trzaskowski, M., Haworth, C. M. A., Davis, O. S. P., Dale, P. S., & Plomin, R. (2012). Socioeconomic status (SES) and children's intelligence (IQ): in a UK-representative sample SES moderates the environmental, not genetic, effect on IQ. *Plos One*, *7*(2), e30320. http://doi.org/10.1371/journal.pone.0030320

Hill, W. D., Hagenaars, S. P., Marioni, R. E., Harris, S. E., Liewald, D. C. M., Davies, G., et al. (2016). Molecular Genetic Contributions to Social Deprivation and Household Income in UK Biobank. *Current Biology*, *26*(22), 3083–3089. http://doi.org/10.1016/j.cub.2016.09.035

Jannot, A.-S., Ehret, G., & Perneger, T. (2015). P < 5 x 10(-8) has emerged as a standard of statistical significance for genome-wide association studies. *Journal of Clinical Epidemiology*, *68*(4), 460–465. http://doi.org/10.1016/j.jclinepi.2015.01.001

Johnson, G. C., Esposito, L., Barratt, B. J., Smith, A. N., Heward, J., Di Genova, G., et al. (2001). Haplotype tagging for the identification of common disease genes. *Nature Genetics*, *29*(2), 233–237. http://doi.org/10.1038/ng1001-233

Knopik, V. S., Neiderhiser, J. M., DeFries, J. C., & Plomin, R. (2017). Behavioural Genetics (7 ed.). London, UK: Worth Publishers.

Korte, A., & Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods*, *9*(29). http://doi.org/10.1186/1746-4811-9-29

Krapohl, E., & Plomin, R. (2016). Genetic link between family socioeconomic status and children's educational achievement estimated from genome-wide SNPs. *Molecular Psychiatry*, *21*(3), 437–443. http://doi.org/10.1038/mp.2015.2

Krapohl, E., Euesden, J., Zabaneh, D., Pingault, J.-B., Rimfeld, K., Stumm, von, S., et al. (2016). Phenome-wide analysis of genome-wide polygenic scores. *Molecular Psychiatry*, *21*(9), 1188–1193. http://doi.org/10.1038/mp.2015.126

Krapohl, E., Hannigan, L. J., Pingault, J.-B., Patel, H., Kadeva, N., Curtis, C., et al. (2017). Widespread covariation of early environmental exposures and trait-associated polygenic variation. *Proceedings of the National Academy of Sciences*, *114*(44), 11727–11732. http://doi.org/10.1073/pnas.1707178114

Lander, E., & Kruglyak, L. (1995). Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nature Genetics*, *11*(3), 241–247. http://doi.org/10.1038/ng1195-241

Lee, J. J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., et al. (2018). Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature Genetics*, *50*(8), 1112–1121. http://doi.org/10.1038/s41588-018-0147-3

Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J., et al. (2009). Finding the missing heritability of complex diseases. *Nature*, *461*(7265), 747–753. http://doi.org/10.1038/nature08494

Marchini, J., & Howie, B. (2010). Genotype imputation for genome-wide association studies. *Nature Reviews. Genetics*, *11*(7), 499–511. http://doi.org/10.1038/nrg2796

Marioni, R. E., Davies, G., Hayward, C., Liewald, D., Kerr, S. M., Campbell, A., et al. (2014). Molecular genetic contributions to socioeconomic status and intelligence. *Intelligence*, *44*(100), 26–32. http://doi.org/10.1016/j.intell.2014.02.006

Mather, K., & Jinks, J. L. (1977). Introduction to biometrical genetics. Cornell Univ Pr.

McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A. R., Teumer, A., et al. (2016). A reference panel of 64,976 haplotypes for genotype imputation. *Nature Genetics*, *48*(10), 1279–1283. http://doi.org/10.1038/ng.3643

Mendel, G. (1866). Versuche über Pflanzenhybriden. *Verhandlungen Des Naturforschenden Vereins in Brünn*, *4*, 3–47.

Nadeau, J. H. (2001). Modifier genes in mice and humans. *Nature Reviews. Genetics*, *2*(3), 165–174. http://doi.org/10.1038/35056009

Neale, B. M., Ferreira, M. A. R., Medland, S. E., & Posthuma, D. (2008). Statistical Genetics: Gene Mapping through Linkage and Association. Abingdon, UK: Taylor and Francis.

Okbay, A., Beauchamp, J. P., Fontana, M. A., Lee, J. J., Pers, T. H., Rietveld, C. A., et al. (2016). Genome-wide association study identifies 74 loci associated with educational attainment. *Nature*, *533*(7604), 539–542. http://doi.org/10.1038/nature17671

Palla, L., & Dudbridge, F. (2015). A Fast Method that Uses Polygenic Scores to Estimate the Variance Explained by Genome-wide Marker Panels and the Proportion of Variants Affecting a Trait. *American Journal of Human Genetics*, *97*(2), 250–259. http://doi.org/10.1016/j.ajhg.2015.06.005

Pardiñas, A. F., Holmans, P., Pocklington, A. J., Escott-Price, V., Ripke, S., Carrera, N., et al. (2018). Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nature Genetics*, *50*(3), 381–389. http://doi.org/10.1038/s41588-018-0059-2

Plomin, R., DeFries, J. C., & Loehlin, J. C. (1977). Genotype-environment interaction and correlation in the analysis of human behavior. *Psychological Bulletin*, *84*(2), 309–322. http://doi.org/10.1037/0033-2909.84.2.309

Polderman, T. J. C., Benyamin, B., de Leeuw, C. A., Sullivan, P. F., van Bochoven, A., Visscher, P. M., & Posthuma, D. (2015). Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nature Genetics*, *47*(7), 702–709. http://doi.org/10.1038/ng.3285

Purcell, S. M., Wray, N. R., Stone, J. L., Visscher, P. M., O'Donovan, M. C., Sullivan, P. F., et al. (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*, *460*(7256), 748–752. http://doi.org/10.1038/nature08185

Rietveld, C. A., Esko, T., Davies, G., Pers, T. H., Turley, P., Benyamin, B., et al. (2014). Common genetic variants associated with cognitive performance identified using the proxy-phenotype method. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(38), 13790–13794. http://doi.org/10.1073/pnas.1404623111

Ripke, S., Sanders, A. R., Kendler, K. S., Levinson, D. F., Sklar, P., Holmans, P. A., et al. (2011). Genome-wide association study identifies five new schizophrenia loci. *Nature Genetics*, *43*(10), 969–976. http://doi.org/10.1038/ng.940

Risch, N., & Merikangas, K. (1996). The future of genetic studies of complex human diseases. *Science*, *273*(5281), 1516–1517. http://doi.org/10.1126/science.273.5281.1516

Rowe, D. C., Jacobson, K. C., & Van den Oord, E. J. (1999). Genetic and environmental influences on vocabulary IQ: parental education level as moderator. *Child Development*, *70*(5), 1151–1162. http://doi.org/10.1111/1467-8624.00084

Rutter, M., Moffitt, T. E., & Caspi, A. (2006). Gene-environment interplay and psychopathology: multiple varieties but real effects. *Journal of Child Psychology and Psychiatry*, *47*(3-4), 226–261. http://doi.org/10.1111/j.1469-7610.2005.01557.x

Schizophrenia Working Group of the Psychiatric Genomics Consortium. (2014). Biological insights from 108 schizophrenia-associated genetic loci. *Nature*, *511*(7510), 421–427. http://doi.org/10.1038/nature13595

Shi, H., Kichaev, G., & Pasaniuc, B. (2016). Contrasting the Genetic Architecture of 30 Complex Traits from Summary Association Data. *The American Journal of Human Genetics*, *99*(1), 139–153. http://doi.org/10.1016/j.ajhg.2016.05.013

Shi, H., Mancuso, N., Spendlove, S., & Pasaniuc, B. (2017). Local Genetic Correlation Gives Insights into the Shared Genetic Architecture of Complex Traits. *American Journal of Human Genetics*, *101*(5), 737–751. http://doi.org/10.1016/j.ajhg.2017.09.022

Sivakumaran, S., Agakov, F., Theodoratou, E., Prendergast, J. G., Zgaga, L., Manolio, T., et al. (2011). Abundant Pleiotropy in Human Complex Diseases and Traits. *The American Journal of Human Genetics*, *89*(5), 607–618. http://doi.org/10.1016/j.ajhg.2011.10.004

Smoller, J. W., Andreassen, O. A., Edenberg, H. J., Faraone, S. V., Glatt, S. J., & Kendler, K. S. (2019). Psychiatric genetics and the structure of psychopathology. *Molecular Psychiatry*, *116*(24), 409–420. http://doi.org/10.1038/s41380-017-0010-4

Solovieff, N., Cotsapas, C., Lee, P. H., Purcell, S. M., & Smoller, J. W. (2013). Pleiotropy in complex traits: challenges and strategies. *Nature Reviews. Genetics*, *14*(7), 483–495. http://doi.org/10.1038/nrg3461

Speed, D., Cai, N., Consortium, T. U., Johnson, M. R., Nejentsev, S., & Balding, D. J. (2017). Reevaluation of SNP heritability in complex human traits. *Nature Genetics*, *49*(7), 986–992. http://doi.org/10.1038/ng.3865

Syvänen, A. C. (2005). Toward genome-wide SNP genotyping. *Nature Genetics*, *37*(6s), S5–S10. http://doi.org/10.1038/ng1558

Teare, M. D., & Barrett, J. H. (2005). Genetic linkage studies. *The Lancet*, *366*(9490), 1036–1044. http://doi.org/10.1016/S0140-6736(05)67382-5

The 1000 Genomes Project Consortium, Gibbs, R. A., Boerwinkle, E., Doddapaneni, H., Han, Y., Korchina, V., et al. (2015). A global reference for human genetic variation. *Nature*, *526*(7571), 68–74. http://doi.org/10.1038/nature15393

The International HapMap Consortium. (2003). The International HapMap Project. *Nature*, *426*(6968), 789–796. http://doi.org/10.1038/nature02168

Trzaskowski, M., Harlaar, N., Arden, R., Krapohl, E., Rimfeld, K., McMillan, A., et al. (2014). Genetic influence on family socioeconomic status and children's intelligence. *Intelligence*, *42*(100), 83–88. http://doi.org/10.1016/j.intell.2013.11.002

Tucker-Drob, E. M., & Bates, T. C. (2016). Large Cross-National Differences in Gene × Socioeconomic Status Interaction on Intelligence. *Psychological Science*, *27*(2), 138–149. http://doi.org/10.1177/0956797615612727

Tucker-Drob, E. M., Rhemtulla, M., Harden, K. P., Turkheimer, E., & Fask, D. (2010). Emergence of a Gene × Socioeconomic Status Interaction on Infant Mental Ability Between 10 Months and 2 Years:. *Psychological Science*, *22*(1), 125–133. http://doi.org/10.1177/0956797610392926

Turkheimer, E. (2000). Three Laws of Behavior Genetics and What They Mean. *Current Directions in Psychological Science*, *9*(5), 160–164. http://doi.org/10.1111/1467-8721.00084

Turkheimer, E., Haley, A., Waldron, M., D'Onofrio, B., & Gottesman, I. I. (2003). Socioeconomic status modifies heritability of IQ in young children. *Psychological Science*, *14*(6), 623–628. http://doi.org/10.1046/j.0956-7976.2003.psci_1475.x

UNESCO Institute for Statistics. (2006). International Standard Classification of Education.

van der Sluis, S., Willemsen, G., de Geus, E. J. C., Boomsma, D. I., & Posthuma, D. (2008). Gene-environment interaction in adults' IQ scores: measures of past and present environment. *Behavior Genetics*, *38*(4), 348–360. http://doi.org/10.1007/s10519-008-9212-5

Vilhjalmsson, B. J., Yang, J., Finucane, H. K., Gusev, A., Lindstrom, S., Ripke, S., et al. (2015). Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores.

American Journal of Human Genetics, *97*(4), 576–592. http://doi.org/10.1016/j.ajhg.2015.09.001

Visscher, P. M., & Yang, J. (2016). A plethora of pleiotropy across complex traits. *Nature Genetics*, *48*(7), 707–708. http://doi.org/10.1038/ng.3604

Visscher, P. M., Brown, M. A., McCarthy, M. I., & Yang, J. (2012a). Five Years of GWAS Discovery. *The American Journal of Human Genetics*, *90*(1), 7–24. http://doi.org/10.1016/j.ajhg.2011.11.029

Visscher, P. M., Goddard, M. E., Derks, E. M., & Wray, N. R. (2012b). Evidence-based psychiatric genetics, AKA the false dichotomy between common and rare variant hypotheses. *Molecular Psychiatry*, *17*(5), 474–485. http://doi.org/10.1038/mp.2011.65

Visscher, P. M., Wray, N. R., Zhang, Q., Sklar, P., McCarthy, M. I., Brown, M. A., & Yang, J. (2017). 10 Years of GWAS Discovery: Biology, Function, and Translation. *The American Journal of Human Genetics*, *101*(1), 5–22. http://doi.org/10.1016/j.ajhg.2017.06.005

Wood, A. R., Esko, T., Yang, J., Vedantam, S., Pers, T. H., Gustafsson, S., et al. (2014). Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature Genetics*, *46*(11), 1173–1186. http://doi.org/10.1038/ng.3097

Wray, N. R., Goddard, M. E., & Visscher, P. M. (2007). Prediction of individual genetic risk to disease from genome-wide association studies. *Genome Research*, *17*(10), 1520–1528. http://doi.org/10.1101/gr.6665407

Wray, N. R., Ripke, S., Mattheisen, M., Trzaskowski, M., Byrne, E. M., Abdellaoui, A., et al. (2018). Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nature Genetics*, *50*(5), 668–681. http://doi.org/10.1038/s41588-018-0090-3

Wray, N. R., Yang, J., Ben J Hayes, Price, A. L., Goddard, M. E., & Visscher, P. M. (2013). Pitfalls of predicting complex traits from SNPs. *Nature Reviews. Genetics*, *14*(7), 507–515. http://doi.org/10.1038/nrg3457

Yang, J., Lee, S. H., Goddard, M. E., & Visscher, P. M. (2011). GCTA: a tool for genome-wide complex trait analysis. *American Journal of Human Genetics*, *88*(1), 76–82. http://doi.org/10.1016/j.ajhg.2010.11.011

Yengo, L., Sidorenko, J., Kemper, K. E., Zheng, Z., Wood, A. R., Weedon, M. N., et al. (2018). Meta-analysis of genome-wide association studies for height and body mass index in similar to 700 000 individuals of European ancestry. *Human Molecular Genetics*, *27*(20), 3641–3649. http://doi.org/10.1093/hmg/ddy271

Chapter 2 – Predicting educational achievement from DNA

This chapter is presented as a published paper. It is an exact copy of this publication:

Selzam, S., Krapohl, E., von Stumm, S., O'Reilly, P. F., Rimfeld, K., Kovas, Y., Dale, P.S., Lee, J.J. & Plomin, R. (2017). Predicting educational achievement from DNA. *Molecular psychiatry*, *22*(2), 267 – 272. doi: 10.1038/mp.2017.203

Supplemental materials are included under Appendix 1.

ORIGINAL ARTICLE

Predicting educational achievement from DNA

S Selzam[1], E Krapohl[1], S von Stumm[2], PF O'Reilly[1], K Rimfeld[1], Y Kovas[1,2,3], PS Dale[4], JJ Lee[5] and R Plomin[1]

A genome-wide polygenic score (GPS), derived from a 2013 genome-wide association study ($N = 127,000$), explained 2% of the variance in total years of education (*EduYears*). In a follow-up study ($N = 329,000$), a new *EduYears* GPS explains up to 4%. Here, we tested the association between this latest *EduYears* GPS and educational achievement scores at ages 7, 12 and 16 in an independent sample of 5825 UK individuals. We found that *EduYears* GPS explained greater amounts of variance in educational achievement over time, up to 9% at age 16, accounting for 15% of the heritable variance. This is the strongest GPS prediction to date for quantitative behavioral traits. Individuals in the highest and lowest GPS septiles differed by a whole school grade at age 16. Furthermore, *EduYears* GPS was associated with general cognitive ability (~3.5%) and family socioeconomic status (~7%). There was no evidence of an interaction between *EduYears* GPS and family socioeconomic status on educational achievement or on general cognitive ability. These results are a harbinger of future widespread use of GPS to predict genetic risk and resilience in the social and behavioral sciences.

Molecular Psychiatry (2017) **22,** 267–272; doi:10.1038/mp.2016.107; published online 19 July 2016

INTRODUCTION

Identifying the genetic variants responsible for the ubiquitous heritability of behavioral dimensions and disorders is transforming genetic research in the social and behavioral sciences by making it possible to predict genetic strengths and weaknesses of individuals from DNA alone.[1] Over the past decade, genome-wide association (GWA) research across the life sciences has revealed that there are almost no genetic variants with large effects on complex traits and common disorders.[2] This consistent finding implies that the heritability of behavioral traits is due to many genetic variants of small effect. GWA studies of behavioral traits began to be successful as their sample sizes increased sufficiently to detect associations of very small effect size between single-nucleotide polymorphisms (SNPs) and outcome.[3] Although the largest effect sizes of the associations between SNPs and behavioral traits are very small, it is possible to aggregate the effects of thousands of SNP associations, ranked by effect size, into a SNP genotypic score for a particular trait.[4–6] Here, we refer to this SNP genotypic score as a genome-wide polygenic score (GPS).[7] Although many different labels have been ascribed to polygenic scores that usually include the word *risk*, we prefer GPS. It highlights the genome-wide nature of these polygenic scores and encompasses positive as well as negative effects implied by the normal distribution of polygenic scores.[4]

The largest GWA analysis of a behaviorally relevant trait so far was performed on years of education, which is a proxy for educational achievement and to a lesser extent for learning ability.[8] Information about the years spent in education is available in many GWA samples because it is a demographic descriptor. In 2013, a GWA analysis of *EduYears* based on 126,559 individuals was published.[9] The corresponding GPS accounted for 2–3% of the variance in years of education in independent samples.[9,10] The latest GWA on years of education published in 2016 included ~329,000 individuals.[8] A revised GPS based on this new GWA almost doubled the effect size, with *EduYears* GPS explaining 3.9% of the variance in years of education in an independent sample.[8]

EduYears GPS has also been associated with other phenotypes, most notably, measured educational achievement. In a Dutch study, the 2013 *EduYears* GPS accounted for around 2% of the variance in educational achievement in a sample of about 1000 children tested at age 12.[11] A UK-based longitudinal study of 4500 participants reported significant associations between the 2013 *EduYears* GPS and educational achievement at 7, 11 and 16;[12] however, the authors did not report the phenotypic variance explained by *EduYears* GPS. In a subsample of the present study of ~3000 individuals, we previously found that the 2013 *EduYears* GPS accounted for about 2% of the variance in educational achievement at age 16.[13]

The present study evaluates the extent to which a GPS constructed on the basis of the published summary statistics of the 2016 GWA analysis of years of education in adulthood predicts educational achievement assessed during the school years, which we have shown to be about 60% heritable estimated by the twin design.[14,15] Using effect size estimates from the 2016 *EduYears* GWA analysis, we calculated a GPS for each individual in a sample of 5825 unrelated UK students for whom we had educational achievement scores at ages 7, 12 and 16 based on UK-wide assessments of the national curriculum.

As mentioned, the 2016 *EduYears* GPS is based on a GWA sample almost three times as large as the 2013 GWA (329,000 versus 127,000), and as a result, the amount of variance that *EduYears* GPS accounted for in the discovery sample doubled (~4 versus 2%). Accordingly, here we tested the extent to which the 2016 *EduYears* GPS accounts for more variance in educational

[1]King's College London, MRC Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology & Neuroscience, London, UK; [2]Department of Psychology, Goldsmiths University of London, London, UK; [3]Laboratory for Cognitive Investigations and Behavioural Genetics, Tomsk State University, Tomsk, Russia; [4]Department of Speech and Hearing Sciences, University of New Mexico, Albuquerque, NM, USA and [5]Department of Psychology, University of Minnesota Twin Cities, Minneapolis, MN, USA. Correspondence: Professor R Plomin, King's College London, MRC Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology & Neuroscience, 16 DeCrespigny Park, London SE5 8AF, UK.
E-mail: robert.plomin@kcl.ac.uk
Received 1 April 2016; revised 10 May 2016; accepted 23 May 2016; published online 19 July 2016

achievement than the 2013 *EduYears* GPS. In addition, we addressed two specific questions about the role of *EduYears* GPS for educational achievement.

First, we tested the extent to which the 2016 *EduYears* GPS is associated with general cognitive ability (*g*, aka intelligence) and with family socioeconomic status (SES), both of which phenotypically correlate with educational achievement ~0.40–0.50.[16] Using summary statistics derived from GWA analyses, a study applying the LD score regression method[17] identified very high genetic correlations between years of education and childhood IQ ($rg = 0.73$).[18] In a subsample of ~3000 individuals from the current study, the 2013 *EduYears* GPS accounted for ~2% of the variability in *g* at age 16.[19] We also reported that this GPS explained ~2.5% of the variance in family SES, which refers to the SES of the children's parents.[13] In the present study, we predicted that the 2016 *EduYears* GPS would yield stronger associations with *g* and family SES than previously found for the 2013 *EduYears* GPS. In addition, we tested whether the 2016 *EduYears* GPS is significantly associated with educational achievement independent of *g* and family SES.

Second, we tested the hypothesis that SES moderates genetic influences on educational achievement and *g*, as predicted by previous studies that observed decreased heritability estimates in low compared with high SES families.[20] This genotype–environment interaction hypothesis leads to the prediction that *EduYears* GPS is more strongly associated with educational achievement and *g* in high compared with low-SES families. In addition, we tested whether this genotype–environment interaction increased from childhood through adolescence as family SES should have a progressively stronger effect on these aspects of children's lives if the genotype–environment interaction hypothesis is correct.

MATERIALS AND METHODS

Participants

This study included unrelated individuals from the multivariate longitudinal Twins Early Development Study that recruited almost 17,000 twin pairs born in England and Wales between 1994 and 1996.[21] The sample is representative of British families in ethnicity, family SES and parental occupation.[21] The genotyped subsample is representative of UK census data at first contact (Supplementary Table S1). The Institute of Psychiatry, Psychology and Neuroscience ethics committee (05.Q0706/228) granted project approval and parental consent was obtained prior to data collection.

DNA for 3497 individuals was extracted from saliva samples and hybridized to HumanOmniExpressExome-8v1.2 genotyping arrays at the MRC SGDP Centre Molecular Genetics Laboratories. The raw image data from the array were normalized, pre-processed, and filtered in GenomeStudio according to Illumina Exome Chip SOP v1.4. (http://confluence.brc.iop.kcl.ac.uk:8090/display/PUB/Production+Version%3A+Illumina+Exome+Chip+SOP+v1.4). In addition, prior to genotype calling, 869 multi-mapping SNPs and 353 samples with call rate < 0.95 were removed. The ZCALL program[22] was used to augment the genotype calling for samples and SNPs that passed the initial QC.

DNA from an additional 3665 samples genotyped earlier in the project was extracted from buccal cheek swabs and genotyped at Affymetrix (Santa Clara, CA, USA). Samples were successfully hybridized to Affymetrix-GeneChip 6.0 SNP genotyping arrays (http://www.affymetrix.com/support/technical/datasheets/genomewide_snp6_datasheet.pdf) using experimental protocols recommended by the manufacturer (Affymetrix). The raw image data from the arrays were normalized and pre-processed at the Wellcome Trust Sanger Institute (Hinxton, UK) for genotyping as part of the Wellcome Trust Case Control Consortium 2 (https://www.wtccc.org.uk/ccc2/) according to the manufacturer's guidelines (http://www.affymetrix.com/support/downloads/manuals/genomewidesnp6_manual.pdf). Genotypes for the Affymetrix arrays were called using CHIAMO (https://mathgen.stats.ox.ac.uk/genetics_software/chiamo/chiamo.html).

After initial quality control and genotype calling, the same quality control was performed on the samples genotyped on the Illumina and Affymetrix arrays separately using PLINK,[23] R[24] and VCFtools.[25] Samples were removed from subsequent analyses on the basis of call rate (< 0.99), suspected non-European ancestry, heterozygosity, array signal intensity (>4 s.d. from the mean) and relatedness. SNPs were excluded if the minor allele frequency was $< 0.05\%$, if more than 1% of genotype data were missing or if the Hardy Weinberg *P*-value was lower than 10^{-5}. Non-autosomal markers and indels were removed. Association between the SNP and the array, batch or plate on which samples were genotyped was calculated; SNPs with an effect *P*-value less than 10^{-3} were excluded. A total sample of 5825 samples, with 2698 individuals genotyped on Illumina and 3127 individuals genotyped on Affymetrix, remained after quality control.

Genotypes from the two arrays were separately imputed using the Haplotype Reference Consortium[26] and Minimac3 1.0.13[27,28] available on the *Michigan Imputation Server* (https://imputationserver.sph.umich.edu) as reference data. A series of quality checks were performed before merging data from the two arrays imputation (e.g. array effects, allele frequencies by imputation quality). For the present analyses, we limited our analyses to variants genotyped or imputed at info > 0.95 on both arrays, and with Hardy Weinberg Equilibrium test *P*-value $> 10^{-5}$. After stringent pruning to remove markers in high linkage disequilibrium ($R^2 > 0.1$) and excluding high linkage disequilibrium genomic regions so as to ensure that only genome-wide effects were detected, we performed Principal Component Analysis on a subset of 40,745 autosomal SNPs that remained after applying our quality control criteria, and that overlapped between the two genotyping arrays. To control for population stratification, we regressed the GPS on the first 10 principal components and used the residuals in all subsequent analyses.

Measures

National Curriculum levels age 7 and 12. English and mathematics National Curriculum levels were collected from teachers when the twins were aged 7 (M = 7.2, s.d. = 0.27) and 12 (M = 11.4, s.d. = 0.66). National Curriculum data and genotypes were available for 4047 children at age 7 and 2950 at age 12. The assessments are based on a rubric aligned with the UK National Curriculum, which is the standardized core academic curriculum formulated by the National Foundation for Educational Research (NFER) and the Qualifications and Curriculum Authority (QCA) (NFER: http://www.nfer.ac.uk/index.cfm; QCA: http://www.qca.org.uk). After receiving parental consent, teachers were contacted directly via mail. Teacher ratings assessed two main abilities: English (including 'speaking and listening', 'reading' and 'writing') and mathematics (including 'using and applying mathematics', 'numbers' and 'shapes, space and measures').

At age 7 and 12, teachers rated National Curriculum levels on a 5-point and 9-point scale, respectively, with higher scores representing greater ability. Mathematics and English abilities correlated 0.74 and 0.81 at age 7 and 12, respectively. Therefore, we created overall academic achievement mean scores by calculating the standardized mean for the English and mathematics scores for both ages.

General Certificate of Secondary Education measures age 16. The General Certificate of Secondary Education (GCSE) is a standardized UK-based examination taken at the end of compulsory education at age 16. In addition to the compulsory core subjects of English, mathematics and science, students can choose from a variety of subjects such as physical education, music, geography, modern foreign languages, and information and communication technology.

GCSE results were obtained by questionnaires sent via mail and by telephone interviews of parents and twins themselves. The grades were coded to range from 4 (G; the minimum pass grade) to 11 (A*; the best possible grade). The GCSE score used in this study represents the mean of the compulsory core subjects mathematics and English (if both English language and English literature were taken, a mean grade for English was derived). The two subjects correlated 0.70. We included only mathematics and English grades in the composite score to improve comparability between the educational achievement measures at the different ages. Self-reported GCSE grades of Twins Early Development Study participants show high accuracy, correlating 0.98 English and 0.99 for mathematics grades with data obtained for a subsample from the National Pupil database (NPD: https://www.gov.uk/government/collections/national-pupil-database).[14] Data for subject grades and genotypes were available for 4301 twins (mean age = 16.62, s.d. = 0.32).

General cognitive ability (g). To measure general cognitive ability, the twins were assessed on various tests including verbal and non-verbal abilities at age 7, 12 and 16. A mean score composite was derived from

four tests ('Conceptual Grouping',[29] 'Similarities',[30] 'Vocabulary',[30] 'Picture Completion'[30]) at age 7; three tests ('Raven's Progressive Matrices',[31] 'General Knowledge'[32] 'Picture Completion'[30]) at age 12; and two tests ('Raven's Progressive Matrices' and 'Mill Hill Vocabulary test') at age 16. Behavioral and genotypic data were available for 3559 individuals at age 7 (M = 7.17, s.d. = 0.29); 3349 individuals at age 12 (M = 11.46, s.d. = 0.64) and 1743 individuals at age 16 (M = 16.52, s.d. = 0.30). General cognitive ability measures at the different ages correlated on average 0.48. For simplicity we created a general cognitive ability mean composite based on data available at ages 7, 12 and 16. Only participants with data from at least two ages were included ($N = 2228$), and mean imputation was performed on those with a missing third measure. We also report results related to general cognitive ability measured at each age individually in Supplementary Table S6.

Family SES. A composite of several factors such as parental education and occupation is considered to reflect SES better than any single factor.[33] Data from 4958 genotyped individuals were available for family SES. This measure represents maternal age at birth of eldest child, the mean score of maternal and paternal highest education level, as well as the respondent's (mother or father) occupation, administered by the Standard Occupational Classification 2000 (Office for National Statistics, 2000) at child age 2, which was the first age of contact.

Small but significant mean differences between girls and boys were found for educational achievement at all ages (Supplementary Table S2). Small age effects were found for educational achievement within each of the three ages (Supplementary Table S2). Therefore, all measures with the exception of SES and *EduYears* GPS were recalculated as standardized residuals corrected for gender and age. To account for a slight negative skew in educational achievement tests at age 7 and 16 and a slight positive skew at age 12, measures were quantile normalized.[34]

Statistical analyses

Genome-wide polygenic scores. We computed GPS for 5825 unrelated individuals using β-weights and *P*-values from summary statistics obtained by GWA analysis. Summary statistics were derived from the 2016 GWA study on years of education[8] with a sample size of 328,918 individuals. It should be noted that the summary statistics we used are slightly different to those of the 2016 *EduYears* study:[8] here 23andMe data are excluded due to legal restrictions, and an initial release of the UK Biobank data are included (see Supplementary Table S3 for cohort details). GPS based on these modified summary statistics correlated highly ($r = 0.86$) with the published GPS[8] when both GPS were constructed using the Health and Retirement Study as target sample. Quality-controlled SNPs were clumped for linkage disequilibrium in PRSice,[35] using $R^2 = 0.1$ cutoff within a 250-kb window. In toal, 108,737 SNPs remained after linkage disequilibrium clumping. We used PRSice[35] to calculate polygenic scores. Firstly, PRSice calculated GPS for each individual in our sample by summing the trait-associated SNPs that are weighted by their effect size derived from GWA analysis. PRSice then performed a regression analysis to test for association between GPS and each of our outcomes (educational achievement at 7, 12, 16, SES and *g*). This is repeated for GPS calculated at a large number of *P*-value thresholds, ranging from 0.001 to 1 (increments of 0.001) in the GWA results, under the high-resolution scoring option in PRSice. Through this high-resolution scoring we identified the 'best-fit' GPS for all measures (Supplementary Table S4), which were used throughout our analyses for each respective trait. The 'best-fit' GPS is identified as that which gives the smallest *P*-value for association with outcome among all the regression tests performed on the GPS (see Supplementary Figures S4). Given the multiple testing involved in high-resolution scoring we use an association significance threshold of $P = 0.001$, as recommended in Euesden *et al.*[35]

For our GPS analyses, we have more than 80% power to explain 0.2% of the phenotypic variance (see Supplementary Methods S1 for details). To test interactions between different levels of *EduYears* GPS and family SES, we have more than 80% power to detect a small interaction effect of $\eta^2 = 0.02$ (given $\alpha = 0.05$; $N = 600$; number of groups = 4).

We performed regression analyses with *EduYears* GPS as a predictor of educational achievement at ages 7, 12 and 16, as well as of *g* and family SES. To test for potential differences between correlations between *EduYears* GPS and educational achievement at the different ages, we performed Fisher's *r*-to-*z* transformations. We also used multiple regression to test whether associations between *EduYears* GPS and educational achievement remain after controlling for family SES and *g*. We also tested for mean differences in educational achievement between the extreme septiles of *EduYears* GPS at each age using analyses of variance. Finally, interaction effects between *EduYears* GPS and SES on educational achievement and on *g* were analyzed using multiple regression models that included each main effect and the interaction effect term.

RESULTS

Polygenic score analyses

As illustrated in Figure 1, *EduYears* GPS accounted for a significant proportion of variance in educational achievement at all ages, increasing from age 7 ($R^2 = 0.028$, $P < 0.001$) to age 12 ($R^2 = 0.046$, $P < 0.001$) to age 16 ($R^2 = 0.091$, $P < 0.001$). Betas indicated that an increase of one standard deviation in *EduYears* GPS resulted in a *z*-standardized mean educational achievement score increase of 0.17, 0.21 and 0.30 at age 7, 12 and 16, respectively. The increase in association between *EduYears* GPS and educational achievement between age 7 and age 16 was significant, as was the association between age 12 and age 16, but not between age 7 and 12 (Supplementary Table S5).

EduYears GPS was also associated with *g* ($R^2 = 0.036$, $P < 0.001$) and family SES ($R^2 = 0.073$, $P < 0.001$) (Figure 1). Additionally, *EduYears* GPS significantly predicted *g* at ages 7, 12 and 16 (Supplementary Table S6); these associations were not statistically different. Because educational achievement, *g*, and family SES are intercorrelated phenotypically (Supplementary Table S6), we tested the effect of *EduYears* GPS on educational achievement independent of *g* and SES by including *g* and SES into a regression model before entering *EduY*ears GPS. After adjusting the *P*-value threshold for multiple testing (see the Materials and methods section), *EduYears* GPS remained a significant predictor of educational achievement at age 16 after accounting for *g* and SES, although the effect size was reduced to 1.2% of the variance explained (Supplementary Table S7).

Extreme group differences

Figure 2 shows the z-standardized mean educational achievement scores by *EduYears* GPS septiles. At all ages, individuals scoring in the highest *EduYears* GPS septile performed on average

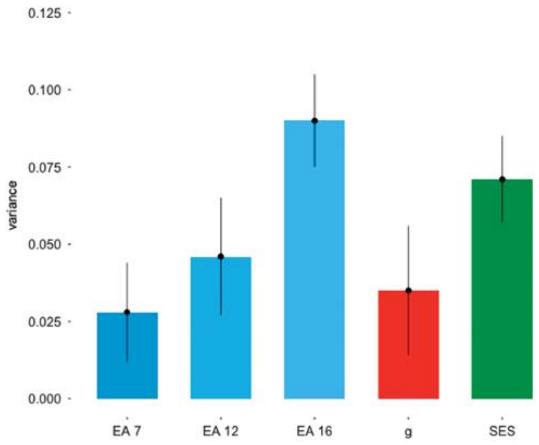


Figure 1. Variance explained (R^2) and standard error of *EduYears* GPS predicting: EA 7 = educational achievement age 7; EA 12 = educational achievement age 12; EA 16 = educational achievement age 16; *g* = general cognitive ability; SES = family socioeconomic status; in this analysis and all subsequent analyses, the unique 'best-fit' GPS was used for each respective trait; see the Materials and methods section for details. GPS, genome-wide polygenic score.

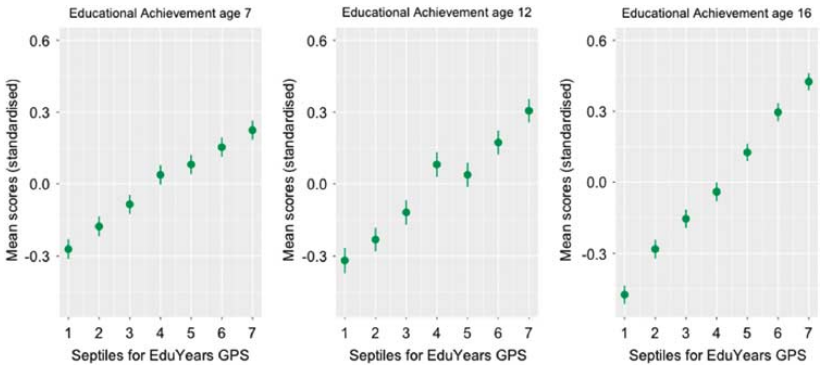


Figure 2. Standardized means and standard errors for educational achievement at age 7, 12 and 16 by genome-wide polygenic score (GPS) septile. *EduYears* GPS was rescored as septiles (1 = lowest, 7 = highest).

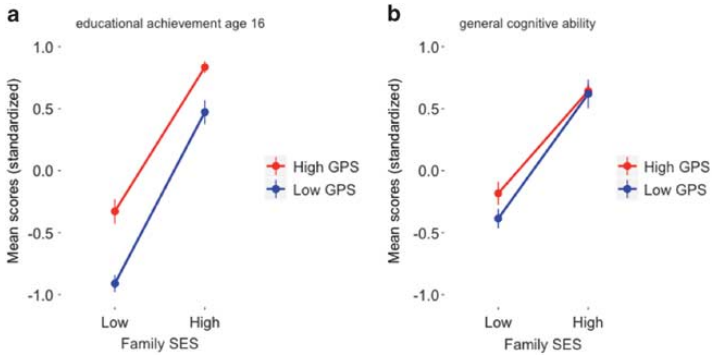


Figure 3. (**a**) Standardized educational achievement mean scores at age 16 by *EduYears* GPS and family SES for individuals scoring in the highest and lowest 20% of the distribution of *EduYears* GPS. There was no evidence for an interaction effect ($F(1,605) = 1.29$, $P = 0.18$); (**b**) general cognitive ability mean scores by *EduYears* GPS and family SES for individuals scoring in the highest and lowest 20% of the distribution of *EduYears* GPS. No interaction effect was found ($F(1,327) = 1.06$, $P = 0.30$). GPS, genome-wide polygenic score; SES, socioeconomic status.

significantly and substantially better at school than those scoring in the lowest GPS septile (Supplementary Table S8). By age 16, there was almost a standard deviation difference in educational achievement between the lowest and highest GPS groups, which represents a whole school grade difference. Similar results were obtained for *EduYears* GPS extreme quintiles rather than septiles (Supplementary Table S9 and Supplementary Figure S1). Using Monte Carlo integration,[36] we calculated a substantial non-overlap of 38% between educational achievement distributions at age 16 for the lowest and highest GPS septiles (Supplementary Figure S2).

Genotype–environment interaction effects

The genetic influence of *EduYears* GPS on educational achievement at age 16 and on *g* was not greater in high SES than in low-SES families, as would be predicted by the genotype–environment interaction hypothesis described earlier. As illustrated in Figure 3a, at age 16 the difference between low and high GPS groups was similar for low-SES and high-SES groups, despite the higher mean educational achievement of the high-SES group. We also did not find G × E interaction for general cognitive ability (Figure 3b), and educational achievement at ages 7 and 12 (Supplementary Figure S3). Hierarchical multiple regression analyses that tested for G × E interaction using continuous data yielded no significant interactions between *EduYears* GPS and SES as they relate to educational achievement at ages 7, 12 and 16 (Supplementary Table S10) or as they relate to *g* (Supplementary Table S11).

DISCUSSION

Our results show that DNA can be used to predict educational achievement, especially at the end of the compulsory school years. Although the 2016 *EduYears* GPS accounted for ~4% of the variance in the GWA target trait of years of education in independent samples, we found that the 2016 *EduYears* GPS accounted for 9% of the variance in educational achievement at age 16, tripling the effect size from previous reports[13] based on the 2013 *EduYears* GPS.[9] The predictive power of *EduYears* GPS can be seen especially at the extremes of the distribution of GPS scores, suggesting that it is possible to identify individuals early in life at genetic risk and resilience, moving us closer to the possibility of early intervention and personalized learning.[37]

We have previously reported a heritability estimate of 60% for educational achievement at age 16 using a sample from which the

present sample was drawn.[14] The present study demonstrated that *EduYears* GPS predicts 9% of the total variance in educational achievement, thus accounting for only 15% of the heritability estimated by the twin design. However, unlike twin study estimates of heritability, GPS is derived from GWA studies, which are limited to additive effects of the common variants employed on SNP arrays. For this reason, SNP-based estimates of heritability, which have these same limitations, represent the current upper limit for GPS prediction. For educational achievement, SNP-based estimates of heritability are about 30%,[13] and *EduYears* GPS explains almost one-third of the heritable variance from SNP-based studies at age 16.

We believe that the substantial increase in heritability explained by the 2016 *EduYears* GPS represents a turning point in the social and behavioral sciences because it makes it possible to predict educational achievement for individuals directly from their DNA. Although other variables account for more of the variance of educational achievement, DNA has a unique predictive status in that inherited DNA sequence variation does not change from the single cell with which life begins. For this reason, unlike the case with many other predictors, the correlation between *EduYears* GPS and educational attainment cannot feasibly be interpreted in terms of reverse causation. That is, the correlation between *EduYears* GPS and educational achievement cannot be caused by the effect of educational achievement on inherited DNA sequence variation. In contrast, although *g* predicts much more of the variance of educational achievement at age 16 (29% in our study), this correlation could be confounded by factors related to both educational achievement and *g*, such as social and family risk factors. Similarly, educational achievement at age 7 predicts 35% of the variance of educational achievement at age 16 but this correlation could also be due to other factors, including genetics,[14] that affect educational achievement at both ages. Moreover, educational achievement and *g* cannot be assessed at earlier stages of development. Family SES, which also predicts substantial variance of educational achievement at age 16 (21% in our study), can be assessed early but this correlation is also likely to be partly caused by other factors, including genetics,[13] that affect both family SES and educational achievement. Although family SES can be assessed early, it can change over time, whereas DNA variations within individuals are stable across the lifespan. Moreover, family SES is a family-wide index not specific to individual children in a family.

EduYears GPS predicts educational achievement independently of *g* and family SES only at age 16, which may be due to the associations between *g,* educational achievement, family SES and *EduYears* GPS. It is possible that family SES and *g* are earlier in the chain of the causal pathway from genetic variants to educational achievement, which may explain the attenuated relationship between *EduYears* GPS and educational achievement at age 7 and 12 after controlling for these variables. Our findings suggest pleiotropic effects of *EduYears* GPS on educational achievement, *g*, and family SES, which are in line with previous reports that describe the genetic overlap between educational achievement, *g,* and family SES.[12,13,38] However, the threefold increase in prediction of educational achievement at age 16 from the 2016 *EduYears* GPS as compared with the 2013 *EduYears* GPS (~3% vs 9%) was not mirrored in the prediction of *g* (~2% vs ~3.5%). The finding that *EduYears* GPS accounts for more variance in educational achievement than in *g* is likely due to the fact that educational achievement is influenced by *g* as well as many other factors that are under genetic influence.[14]

Variance explained by the 2016 *EduYears* GPS in family SES also increased almost threefold compared with previous results with the 2013 EduYears GPS in the a subsample of the current study (~2.5% vs ~7%).[13] Explaining ~7% in family SES by *EduYears* GPS is impressive for two reasons. First, the children's genotypes are only an approximation of their parents' genotypes; the effect of *EduYears* GPS on SES should be even stronger for the parents' own GPS. Second, our findings account for a third of the SNP-based heritability estimate for family SES (~20%),[39] which, as noted earlier, represents the upper limit for GWA and GPS studies. With that, our results demonstrate that family SES is genetically influenced and that its genetic effects are also partly shared with educational achievement.

When interpreting the current results, three caveats should be considered. First, the finding that the predictive validity of *EduYears* GPS increases across the school years may be due to increasing approximation of our measures to the *EduYears* GWA target trait of years of education. That is, our measure of educational achievement at age 16 is a standardized examination taken at the end of compulsory education that strongly influences whether pupils go on to higher education. Alternatively, it is also possible that GCSE results are more reliable measures than national curriculum teacher ratings, which might contribute to the difference in variance explained in these variables by *EduYears* GPS. Second, as we measured family SES in a traditional way by including parental education, this could have increased the association of the SES composite with *EduYears* GPS. Although parental education and occupation are related, future studies should investigate if the relationship between *EduYears* GPS and SES varies as a function of different SES indicators. Third, our finding that *EduYears* accounts for 9% of the variance of educational achievement at age 16 needs to be tested for generalization in other samples and beyond the UK.

The finding that individuals' polygenic scores for years of education predict educational achievement entails no necessary policy implications. However, our findings corroborate that individual differences in educational achievement are partly due to DNA differences between children and are not solely created by environmental forces. By creating a dialogue between scientists and policymakers, the introduction of polygenic scores may soon become a useful tool for early prediction and prevention of educational problems and for personalized learning.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

We gratefully acknowledge the ongoing contribution of the participants in the Twins Early Development Study (TEDS) and their families. TEDS is supported by a program grant to RP from the UK Medical Research Council (MR/M021475/1 and previously G0901245), with additional support from the US National Institutes of Health (HD044454; HD059215). SS is supported by the MRC/IoPPN Excellence Award and by the EU Framework Programme 7 (602768). EK and KR are supported by a Medical Research Council studentship. RP is supported by a Medical Research Council Advanced Investigator award (295366). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

AUTHOR CONTRIBUTIONS

RP directs and received funding for the Twins Early Development Study (TEDS). RP and SS conceived the present study. SS analyzed and interpreted the data. RP supervised the project and interpreted the data. RP and SS wrote the manuscript with help from EK, SvS, PFO, KR, YK, PSD and JJL.

REFERENCES

1 Plomin R, Simpson MA. The future of genomics for developmentalists. *Dev Psychopathol* 2013; **25**: 1263–1278.
2 Chabris CF, Lee JJ, Cesarini D, Benjamin DJ, Laibson DI. The fourth law of behavior genetics. *Curr Dir Psychol Sci* 2015; **24**: 304–312.
3 Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. *Am J Hum Genet* 2012; **90**: 7–24.

4 Dudbridge F. Power and predictive accuracy of polygenic risk scores. *PLoS Genet* 2013; **9**: e1003348.
5 Harlaar N, Butcher LM, Meaburn E, Sham P, Craig IW, Plomin R. A behavioural genomic analysis of DNA markers associated with general cognitive ability in 7-year-olds. *J Child Psychol Psychiatry* 2005; **46**: 1097–1107.
6 Wray NR, Lee SH, Mehta D, Vinkhuyzen AAE, Dudbridge F, Middeldorp CM. Research review: polygenic methods and their application to psychiatric traits. *J Child Psychol Psychiatry* 2014; **55**: 1068–1087.
7 Plomin R, Deary IJ. Genetics and intelligence differences: five special findings. *Mol Psychiatry* 2015; **20**: 98–108.
8 Okbay A, Beauchamp JP, Fontana MA, Lee JJ, Pers TH, Rietveld CA *et al.* Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* 2016; **533**: 539–542.
9 Rietveld CA, Medland SE, Derringer J, Yang J, Esko T, Martin NW *et al.* GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science* 2013; **340**: 1467–1471.
10 Domingue BW, Belsky DW, Conley D, Harris KM, Boardman JD. Polygenic influence on educational attainment. *AERA Open* 2015; **1**: 2332858415599972.
11 de Zeeuw EL, van Beijsterveldt CEM, Glasner TJ, Bartels M, Ehli EA, Davies GE *et al.* Polygenic scores associated with educational attainment in adults predict educational achievement and ADHD symptoms in children. *Am J Med Gen Part B* 2014; **165**: 510–520.
12 Davies NM, Hemani G, Timpson NJ, Windmeijer F, Davey Smith G. The role of common genetic variation in educational attainment and income: evidence from the National Child Development Study. *Sci Rep* 2015; **5**: 16509.
13 Krapohl E, Plomin R. Genetic link between family socioeconomic status and children's educational achievement estimated from genome-wide SNPs. *Mol Psychiatry* 2015; **21**: 437–443.
14 Krapohl E, Rimfeld K, Shakeshaft NG, Trzaskowski M, McMillan A, Pingault J-B *et al.* The high heritability of educational achievement reflects many genetically influenced traits, not just intelligence. *Proc Natl Acad Sci USA* 2014; **111**: 15273–15278.
15 Kovas Y, Haworth CMA, Dale PS, Plomin R. The genetic and environmental origins of learning abilities and disabilities in the early school years. *Monogr Soc Res Child Dev* 2007; **72**: vii, 1–144.
16 Strenze T. Intelligence and socioeconomic success: a meta-analytic review of longitudinal research. *Intelligence* 2007; **35**: 401–426.
17 Bulik-Sullivan BK, Loh P-R, Finucane HK, Ripke S, Yang J. Schizophrenia Working Group of the Psychiatric Genomics *et al.* LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* 2015; **47**: 291–295.
18 Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh P-R *et al.* An atlas of genetic correlations across human diseases and traits. *Nat Genet* 2015; **47**: 1236–1241.
19 Krapohl E, Euesden J, Zabaneh D, Pingault JB, Rimfeld K, von Stumm S *et al.* Phenome-wide analysis of genome-wide polygenic scores. *Mol Psychiatry* 2015; e-pub ahead of print 25 August 2015.
20 Tucker-Drob EM, Bates TC. Large cross-national differences in gene × socioeconomic status interaction on intelligence. *Psychol Sci* 2015; **27**: 138–149.
21 Haworth C, Davis OS, Plomin R. Twins Early Development Study (TEDS): a genetically sensitive investigation of cognitive and behavioral development from childhood to young adulthood. *Twin Res Hum Genet* 2013; **16**: 117–125.
22 Goldstein JI, Crenshaw A, Carey J, Grant GB, Maguire J, Fromer M *et al.* zCall: a rare variant caller for array-based genotyping. *Bioinformatics* 2012; **28**: 2543–2545.
23 Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007; **81**: 559–575.
24 R Core Team . *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing [Internet], 2013. Available at https://www.r-project.org.
25 Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA *et al.* The variant call format and VCFtools. *Bioinformatics* 2011; **27**: 2156–2158.
26 McCarthy S, Das S, Kretzschmar W, Durbin R, Abecasis G, Marchini J. A reference panel of 64,976 haplotypes for genotype imputation. *bioRxiv* 2015; 035170.
27 Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet* 2012; **44**: 955–959.
28 Fuchsberger C, Abecasis GR, Hinds DA. minimac2: faster genotype imputation. *Bioinforma Oxf Engl* 2015; **31**: 782–784.
29 McCarthy D. *McCarthy Scales of Children's Abilities*. The Psychological Corporation: New York, 1972.
30 Wechsler D. *Wechsler Intelligence Scale for Children*, 3rd edn. The Psychological Corporation: UK, 1992.
31 Raven JC, Court JH, Raven J. *Manual for Raven's Progressive Matrices and Vocabulary Scales*. Oxford University: Oxford, 1996.
32 Kaplan E, Fein D, Kramer J, Delis D, Morris R. *WISC-III as a Process Instrument (WISC-III-PI)*. The Psychological Corporation: New York, 1999.
33 White KR. The relation between socioeconomic status and academic achievement. *Psychol Bull* 1982; **91**: 461–481.
34 Van der Waerden BL. On the sources of my book Moderne Algebra. *Hist Math* 1975; **2**: 31–40.
35 Euesden J, Lewis CM, O'Reilly PF. PRSice: Polygenic Risk Score software. *Bioinformatics* 2015; **31**: 1466–1468.
36 Hammersley JM, Handscomb DC. *Monte Carlo Methods*. Wiley: Hoboken, 1964.
37 Asbury K, Plomin R. *G is for Genes: The Impact of Genetics on Education and Achievement*, 1st edn. John Wiley & Sons: Chichester, West Sussex, 2013.
38 Marioni RE, Davies G, Hayward C, Liewald D, Kerr SM, Campbell A *et al.* Molecular genetic contributions to socioeconomic status and intelligence. *Intelligence* 2014; **44**: 26–32.
39 Hanscombe KB, Trzaskowski M, Haworth CMA, Davis OSP, Dale PS, Plomin R. Socioeconomic status (SES) and children's intelligence (IQ): in a UK-representative sample SES moderates the environmental, not genetic, effect on IQ. *PLoS One* 2012; **7**: e30320.

Supplementary Information accompanies the paper on the Molecular Psychiatry website (http://www.nature.com/mp)

Chapter 3 – Polygenic score for educational attainment captures DNA variants shared between personality traits and educational achievement

This chapter is presented as a published paper. It is an exact copy of this publication:

Smith-Woolley, E.*, **Selzam, S.***, & Plomin, R. (2019). Polygenic Score for Educational Attainment Captures DNA Variants Shared Between Personality Traits and Educational Achievement. *Journal of Personality and Social Psychology: Personality Processes and Individual Differences.* Advanced online publication. doi: 10.1037/pspp0000241

*Joint first authors

Supplemental materials are included under Appendix 2.

Journal of Personality and Social Psychology:
Personality Processes and Individual Differences

0022-3514/19/$12.00

http://dx.doi.org/10.1037/pspp0000241

Polygenic Score for Educational Attainment Captures DNA Variants Shared Between Personality Traits and Educational Achievement

Emily Smith-Woolley, Saskia Selzam, and Robert Plomin
King's College London

Genome-wide polygenic scores (GPS) can be used to predict individual genetic risk and resilience. For example, a GPS for years of education (*EduYears*) explains substantial variance in cognitive traits such as general cognitive ability and educational achievement. Personality traits are also known to contribute to individual differences in educational achievement. However, the association between *EduYears* GPS and personality traits remains largely unexplored. Here, we test the relation between GPS for *EduYears*, neuroticism, and well-being, and 6 personality and motivation domains: Academic Motivation, Extraversion, Openness, Conscientiousness, Neuroticism, and Agreeableness. The sample was drawn from a U.K.-representative sample of up to 8,322 individuals assessed at age 16. We find that *EduYears* GPS was positively associated with Openness, Conscientiousness, Agreeableness, and Academic Motivation, predicting between 0.6% and 3% of the variance. In addition, we find that *EduYears* GPS explains between 8% and 16% of the association between personality domains and educational achievement at the end of compulsory education. In contrast, both the neuroticism and well-being GPS significantly accounted for between 0.3% and 0.7% of the variance in a subset of personality domains. Furthermore, they did not significantly account for any of the covariance between the personality domains and achievement, with the exception of the neuroticism GPS explaining 5% of the covariance between Neuroticism and achievement. These results demonstrate that the genetic effects of educational attainment relate to personality traits, highlighting the multifaceted nature of *EduYears* GPS.

Keywords: academic achievement, motivation, personality, polygenic score

Supplemental materials: http://dx.doi.org/10.1037/pspp0000241.supp

Education is one of society's most expensive intervention programs. Among the member countries of the Organisation for Economic Cooperation and Development (OECD), education accounts for between 6 and 15% of annual gross domestic product (OECD, 2017) and the average young person in these countries will stay in education until the age of 22 (OECD, 2007). Given its societal value, great importance is placed on succeeding in education, both in terms of educational attainment (education level) and education achievement (education grade).

For a century, psychologists have attempted to unravel the major predictors of individual differences in educational success. Early work showed that *cognitive capacity* played a substantial role in education performance (Binet & Simon, 1916), a term that now many refer to as general cognitive ability or *g*. However, it did not

Emily Smith-Woolley, Saskia Selzam, and Robert Plomin, MRC Social, Genetic, and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology, and Neuroscience, King's College London.

Emily Smith-Woolley is now at the Centre for Education Improvement Science, UCL Institute of Education, University College London, UK.

Emily Smith-Woolley and Saskia Selzam are joint first authors.

Author Contributions. Study concept and design: Emily Smith-Woolley, Saskia Selzam, Robert Plomin. Processed and quality controlled genotype data: Saskia Selzam. Analysis of data: Emily Smith-Woolley, Saskia Selzam. Interpretation of data: Emily Smith-Woolley, Saskia Selzam, Robert Plomin. Wrote the paper: Emily Smith-Woolley, Saskia Selzam, Robert Plomin.

We gratefully acknowledge the ongoing contribution of the participants in the Twins Early Development Study (TEDS) and their families. TEDS is supported by a program grant to RP from the UK Medical Research Council (MR/M021475/1 and previously G0901245), with additional support from the US National Institutes of Health (AG046938). The research leading to these results has also received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007–2013)/ grant agreement n° 602768 and ERC grant agreement n° 295366. RP is supported by a Medical Research Council Professorship award (G19/2). SS is supported by the MRC/IoPPN Excellence Award and by the U.S. National Institutes of Health (AG046938). High performance computing facilities were funded with capital equipment grants from the GSTT Charity (TR130505) and Maudsley Charity (980).

Correspondence concerning this article should be addressed to Emily Smith-Woolley, MRC Social, Genetic, and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology, and Neuroscience, King's College London, Memory Lane, Camberwell, London SE5 8AF, United Kingdom. E-mail: emily.smith-woolley@kcl.ac.uk

tell the whole story. Around the same time, Webb (1915) proposed that in addition to g, academic performance was also influenced by a *w* or *will* factor, representing drive or motivation (Webb, 1915). This led the way for psychological explanations of educational success. Most now accept a more complex model of academic performance that comprises both what a person can do (general cognitive ability) and how a person will do it (personality, motivation and other psychosocial influences).

One important factor, influencing both the can and the how, is genetics. Inherited DNA differences play an important role in explaining individual differences in personality traits, general cognitive ability and educational outcomes. Decades of research using twin studies have shown substantial heritability for personality traits, general cognitive ability and educational outcomes (Polderman et al., 2015). To estimate genetic and environmental influences using twin studies, the relative similarities between identical (monozygotic [MZ]) twins, who share 100% of their inherited DNA, are compared with the relative similarities between fraternal (dizygotic [DZ]) twins, who share on average 50% of their inherited DNA differences (Knopik, Neiderhiser, DeFries, & Plomin, 2017). Because both sets of twins grow up in equally similar environments (Derks, Dolan, & Boomsma, 2006; Kendler, Neale, Kessler, Heath, & Eaves, 1993), the influence of genetics and the environment on traits can be unpacked: if MZ twins correlate higher for a trait than DZ twins, then genetic influence is inferred. However, twin studies can only tell us about the relative genetic influence on differences in traits within a population, rather than the influence of measured DNA differences on traits. In the current study, we use a more recent, genetically sensitive method—genome-wide polygenic scoring—to predict a broad range of personality and motivation traits directly from DNA. Furthermore, we estimate the role of measured DNA in the association between these personality traits and academic achievement at age 16.

General Cognitive Ability and Educational Performance

Educational achievement represents a cumulative process of acquiring many skills, gradually over time. Although it is influenced by a multitude of different factors, one of the most powerful and parsimonious predictors of educational achievement is general cognitive ability. General cognitive ability captures the communalities within a diverse set of cognitive measures, such as memory, verbal-reasoning, and nonverbal reasoning (Plomin & Deary, 2015). It is highly correlated with academic achievement at age 9 ($r = .45$; Spinath, Spinath, Harlaar, & Plomin, 2006), school performance at the end of compulsory education at age 16 ($r = .81$; Deary, Strand, Smith, & Fernandes, 2007), and also later with university achievement ($r = .48$; Frey & Detterman, 2004).

In addition to educational achievement, general cognitive ability is also strongly associated with years spent in full-time education (Deary & Johnson, 2010; Jencks, 1979; Ritchie & Tucker-Drob, 2018). However, although general cognitive ability explains more than half of the variance in academic outcomes (Deary et al., 2007), it still leaves a substantial portion of the variance unexplained. Therefore, it is important to consider other explanatory factors influencing educational performance.

Personality and Educational Performance

The most widely researched personality correlates of educational performance are dimensions of the Five-Factor Model (FFM; McCrae & Costa, 1987). The FFM comprises Conscientiousness (dependability and drive to achieve), Extraversion (sociability and activity), Openness to Experience (curiosity and broadmindedness), Agreeableness (compassion and kindness), and Neuroticism (stress and anxiety). These broad domains have been linked both positively (conscientiousness, openness and agreeableness) and negatively (neuroticism and extraversion) to academic performance (Busato, Prins, Elshout, & Hamaker, 2000; Chamorro-Premuzic & Furnham, 2003; Conard, 2006; De Raad & Schouwenburg, 1996; O'Connor & Paunonen, 2007; Petrides, Chamorro-Premuzic, Frederickson, & Furnham, 2005; Poropat, 2009; Richardson, Abraham, & Bond, 2012). In addition, their underlying, specific facets (most notably dutifulness, achievement-striving and anxiety) have also been associated with differences in academic performance (Chamorro-Premuzic & Furnham, 2003).

Many studies have explored the reasons for observed associations between FFM dimensions and academic performance—both in terms of attainment and achievement. Conscientiousness is comparable with the *w* factor described by Webb (1915) and has been linked to academic effort (Trautwein, Lüdtke, Roberts, Schnyder, & Niggli, 2009) through time spent on homework (Trautwein & Lüdtke, 2007) and time use efficiency (Kelly & Johnson, 2005). It has been shown to predict academic performance at high-school (Heaven & Ciarrochi, 2008; Laidra, Pullmann, & Allik, 2007), undergraduate (Chamorro-Premuzic & Furnham, 2003; Conard, 2006; Wagerman & Funder, 2007) and even at postgraduate level (Hirschberg & Itkin, 1978). Agreeableness and Openness have also been linked to academic performance: Agreeableness through following teacher instructions and learning style (Busato, Prins, Elshout, & Hamaker, 1999), and Openness through critical thinking (Bidjerano & Dai, 2007) and intelligence (Holland, Dollinger, Holland, & Macdonald, 1995; McCrae & Costa, 1997). Like Conscientiousness, Openness is also related to success in school and at university, showing positive correlations with undergraduate and postgraduate examination scores (Geramian, Mashayekhi, & Ninggal, 2012; Laidra et al., 2007). In contrast, Neuroticism and Extraversion have been negatively linked to academic achievement; Extraversion through distractibility, sociability, and problems regulating effort devoted to academic tasks (Bidjerano & Dai, 2007), and Neuroticism through stress linked with exams and poor impulse control (Zeidner & Matthews, 2000).

Because there are intercorrelations between personality traits, general cognitive ability, and academic achievement, an important question to consider is how these personality traits link to achievement beyond cognitive ability. Conscientiousness has consistently been linked to academic achievement over and above general cognitive ability. For example, it was demonstrated (Poropat, 2009) that Conscientiousness was largely independent of intelligence and that when academic achievement at high school was accounted for, Conscientiousness continued to predict achievement at university. This is in line with another study also showing that once prior achievement on SATs was accounted for, Conscientiousness incrementally predicted later achievement (Conard, 2006). However, there have been few studies looking at person-

ality and general cognitive ability concurrently at secondary school level.

Motivation and Educational Performance

In addition to personality dimensions, other explanations of academic performance have been put forward. In a systematic review of psychological traits, Richardson and colleagues (Richardson et al., 2012) suggest five nonintellective domains influencing educational success: (a) personality traits, (b) motivational factors, (c) self-regulatory strategies, (d) student's approaches to learning, and (e) psychosocial influences. Although the authors note that these domains are conceptually overlapping, they argue that it is important to consider a wide variety of nonintellective factors when predicting academic performance.

One of these factors, which has consistently been linked to academic performance, is motivation. Although aspects of motivation correlate moderately with the FFM dimensions, for example Extraversion (positively) and Neuroticism (negatively; Komarraju & Karau, 2005), many argue that elements of motivation, such as self-efficacy beliefs, may influence achievement over and above these dimensions (Caprara, Vecchione, Alessandri, Gerbino, & Barbaranelli, 2011).

Self-efficacy beliefs are an individual's beliefs about their capabilities to produce effects (Bandura, 1997). Self-efficacy and related traits, such as self-perceived ability, engagement, and academic self-concept are important constructs which help to explain students' learning and progress (Multon, Brown, & Lent, 1991; Schunk, 1989). In one study specifically looking at math self-efficacy and self-concept (Parker, Marsh, Ciarrochi, Marshall, & Abduljabbar, 2014), moderate correlations with achievement in math and science were found (r = .17–.58), and math self-efficacy was also a significant predictor of university entry. Similarly to personality dimensions, self-efficacy beliefs have also been shown to predict academic achievement over and above general cognitive ability; self-perceptions of ability explained an extra 8% of the variance in math achievement and 9% in English achievement at age nine after accounting for general cognitive ability (Spinath et al., 2006).

Heritability of Personality Traits

The heritability of personality traits has been well established. Estimates of the genetic influence on variance in the Big Five personality traits range from 40–60% (Bouchard, Jr. & McGue, 2003; Jang, Livesley, & Vemon, 1996; Polderman et al., 2015). In line with twin study heritability estimates of personality traits, one twin study using the same sample as in the present study found that, at age 16, heritability ranged from 35% for well-being to 40% for self-efficacy and up to 46% for aspects of personality (Krapohl et al., 2014). Furthermore, in the same study, they found that inherited DNA differences explained a large portion of the observed correlation between personality and general cognitive ability and academic achievement. Consistent with this, a study using twins from the United States also found that genetically influenced variation accounted for the associations between personality traits and both academic achievement and verbal knowledge (Tucker-Drob, Briley, Engelhardt, Mann, & Harden, 2016). In addition, they found that part of these genetically mediated associations were shared with general cognitive ability. This suggests that some of the genetic factors driving variation in personality and general cognitive ability are also explaining variance in achievement. This concept is known as pleiotropy—the finding that single genetic variants affect multiple traits (Solovieff, Cotsapas, Lee, Purcell, & Smoller, 2013).

Although twin studies are not able to point to specific genetic variants that are responsible for covariation between traits, the extent to which the phenotypic correlation between traits can be explained by genetics (the genetic correlation) is an index of pleiotropy. Why might genetic variants associated with personality and general cognitive ability also be related to achievement? Doing well in exams requires more than just intelligence; it requires motivation, concentration, diligence, good mental health, as well as many other factors. Furthermore, these heritable traits might also lead individuals to choose certain environments for themselves, for example, individuals high on Conscientiousness may choose to attend optional revision classes and complete homework on time. These decisions may in turn lead to better educational outcomes, such as higher grades. This illustrates a concept known as gene-environment correlation (rGE; Knopik et al., 2017; Plomin, DeFries, & Loehlin, 1977). rGE is the idea that an individual's genetically influenced behavior may elicit specific reactions from others (evocative rGE), or lead individuals to choose experiences and environments that correlate with their genotype (active rGE). A third type of rGE is passive rGE, whereby children are exposed to family environments that are partly created by, and therefore correlated with, their parents' genetic propensities. If passive rGE is at play, these inherited environments reinforce children's own genetic propensities, driving development, or codevelopment of traits. Indeed, recent studies have shown that passive rGE is a likely mechanism in the development of educational achievement (Kong et al., 2018; Lee et al., 2018). Presented in this context, finding that much of the correlation between personality and educational achievement is explained by genetic factors, may therefore be partly reflecting a developmental pattern induced by rGE.

Using DNA to Predict Personality Traits

In addition to family studies, such as twin designs, DNA-based methods have also shed light on genetic influence on personality traits. Genome-wide association (GWA) studies test associations between millions of known DNA variants, called single nucleotide polymorphisms (SNPs), and phenotypic traits in large samples comprising thousands of individuals. GWA studies have shown that effect sizes between individual SNPs and complex traits are usually very small, with single SNPs generally explaining less than 0.1% of the variance each (Gratten, Wray, Keller, & Visscher, 2014). However, because it is assumed that most of these genetic effects are additive, more phenotypic variance can be explained when considering these SNPs jointly (Purcell et al., 2009). By summing up the number of trait-increasing alleles, which are weighted by the GWA SNP effect sizes across thousands of SNPs, it is possible to generate a genetic score for each individual in an independent sample. These genetic scores, referred to as genome-wide polygenic scores (GPS), allow DNA-based prediction for any complex trait.

One of the largest published GWA studies for a behavioral trait is years of education (*EduYears*; Lee et al., 2018; Okbay, Baselmans, et al., 2016; Rietveld et al., 2013). This study, which had a sample size of 1.1 million adults, tested associations between

SNPs and total years in education. It is possible to use the results from this study, indicating which SNPs are associated with years of education and how large the association is, to create GPS in an independent, genotyped sample. *EduYears* GPS has been shown to explain 11–13% of the variance in the target trait years of education (Lee et al., 2018), 7–10% in cognitive performance (Lee et al., 2018), up to 5% in reading ability (Selzam, Dale, et al., 2017), and up to 15% in educational achievement at 16 (Allegrini et al., 2018).

Although cognitive GPS such as *EduYears* and intelligence appear to be explaining variance in their target traits, and related traits such as achievement (Plomin & von Stumm, 2018), personality GPS have been less predictive. For example, a GPS for well-being explains 0.9% of the variance in well-being and 0.7% in neuroticism (Okbay, Baselmans, et al., 2016). In the current study, we sought to investigate whether *EduYears* GPS could predict variance in a range of personality and motivation domains, how this prediction compared with personality polygenic score prediction, and whether personality polygenic scores relate to educational achievement.

Why might a genome-wide polygenic score for education link to personality? Similarly to achievement, educational attainment (years in education) is influenced by a multitude of heritable traits in both the cognitive ability and personality domains (Fredricks, Blumenfeld, & Paris, 2004). So far, only one study (Mõttus, Realo, Vainik, Allik, & Esko, 2017) has related *EduYears* GPS to personality traits. This study investigated the link between *EduYears* GPS and the Big Five personality traits in an Estonian sample of ~3,000 adults of a wide age range. *EduYears* GPS predicted 0.5% of the variance in Neuroticism and 1.2% in Openness to experience, suggesting that the polygenic score for educational attainment tags genetic variants that also relate to personality domains. However so far, no study has investigated links to other personality traits aspects, such as the underlying, more specific facets of personality (e.g., well-being or anxiety), as well as motivation traits such as self-efficacy beliefs.

The Present Study

Given the genetic links between personality traits and educational achievement, the current study sought to explore these associations further by testing the extent to which *EduYears* GPS correlated with personality and motivation domains, as well as their subtraits. In addition, using a neuroticism GPS and well-being GPS, we contrasted the association between these personality GPS and educational achievement to *EduYears* GPS. We also tested whether associations remained after accounting for general cognitive ability. Finally, given previous quantitative genetics findings, we tested the extent to which the *EduYears*, neuroticism and well-being GPS explain the covariance between a range of personality traits and educational achievement at age 16.

Method

Ethics

Ethical approval for this study was received from King's College London Ethics Committee, Reference Number: PNM/09/10–104.

Sample

The sampling frame for the present study was the Twins Early Development Study (TEDS; Haworth, Davis, & Plomin, 2013). TEDS includes 16,000 twin pairs born between 1994 and 1996 and followed from birth to the present day. Although there has been some attrition, approximately 10,000 twin pairs are still enrolled in the study, providing behavioral, cognitive, and psychological data. The TEDS sample is representative of families with children in England and Wales (Haworth et al., 2013). The current study uses a genotyped subsample of TEDS which comprises 10,346 Caucasian individuals, including 7,026 unrelated individuals (i.e., one member of a twin pair) and 3,320 DZ cotwins. Written informed consent was obtained from parents before data collection.

Genotyping

Two genotyping platforms were used to genotype TEDS individuals because these genotyping efforts were separated by 5 years. AffymetrixGeneChip 6.0 SNP arrays were used to genotype 3,747 individuals at Affymetrix, Santa Clara (CA) based on buccal cell DNA samples. Genotypes were generated at the Wellcome Trust Sanger Institute (Hinxton, U.K.) as part of the Wellcome Trust Case Control Consortium 2 (https://www.wtccc.org.uk/ccc2/). Additionally, 8,122 individuals, including 3,607 dizygotic twin pairs, were genotyped on HumanOmniExpressExome-8v1.2 arrays at the Molecular Genetics Laboratories of the Medical Research Council Social, Genetic and Developmental Psychiatry Centre, based on DNA that was extracted from saliva samples. A total sample of 10,346 samples (including 3,320 dizygotic twin pairs and 7,026 unrelated individuals), with 7,289 individuals and 559,772 SNPs genotyped on Illumina and 3,057 individuals and 635,269 SNPs genotyped on Affymetrix remained after quality control. Both samples were imputed separately to the Haplotype Reference Consortium (release 1.1) reference genotypes using the Sanger Imputation Server (McCarthy et al., 2016), before merging genotype data obtained from both platforms. Following postimputation quality control and platform harmonization, 7,363,646 SNPs were retained for the analyses (for full details, see Selzam et al., 2018).

To calculate genomic principal components to account for population stratification, we performed principal component analysis on a subset of 39,353 common (MAF $>$ 5%), perfectly imputed (info = 1) autosomal SNPs, after stringent pruning to remove markers in linkage disequilibrium ($r^2 > 0.1$) and exclusion of high linkage disequilibrium genomic regions.

Measures

General Certificate of Secondary Education. The General Certificate of Secondary Education (GCSE) is a standardized U.K.-based examination at the end of compulsory education at age 16. Students are required to take three core subjects: English, mathematics, and science. For 7,325 genotyped individuals, these results were obtained from questionnaires sent via mail, in addition to telephone interviews with twins and their parents. We also obtained subject grades for an additional 1,227 geno-

typed participants who had missing self-reported data from the National Pupil database (NPD: https://www.gov.uk/government/collections/national-pupil-database). Written consent was given before accessing these data. The total sample included 8,552 genotyped individuals (M = 16.30 years; SD = 0.29 years), including 2,799 DZ twin pairs. Subjects were graded from 4 (G; the minimum pass grade) to 11 (A*; the best possible grade). We used a mean of the three z-standardized compulsory subjects because other subjects are taken by only subsamples of the students. English, mathematics and science performance correlated highly with each other (r = .70–.81). Furthermore, self-reported GCSE grades of TEDS participants show high accuracy, correlating 0.98 English and 0.99 for mathematics grades with data obtained for a subsample from the NPD.

General cognitive ability. Individuals were measured on multiple cognitive tests including verbal and nonverbal abilities at age 7 (M = 7.12, SD = 0.24, N = 5,612), 12 (M = 11.44, SD = 0.65, N = 5,284), and 16 (M = 16.47, SD = 0.278, N = 2,840). Age specific mean score composites were derived from four tests at age seven: Conceptual Grouping (McCarthy, 1972), Similarities, Vocabulary, and Picture Completion (Wechsler, Golombok, & Rust, 1992); three tests at age 12: Raven's Progressive Matrices (Raven, Raven, & Court, 1998), General Knowledge (Kaplan, Fein, Kramer, Delis, & Morris, 1999), and Picture Completion (Wechsler et al., 1992) and two tests at age 16: Raven's Progressive Matrices (Raven et al., 1998) and Mill Hill Vocabulary test (Raven, Raven, & Court, 1989). A general cognitive ability composite was created by taking the arithmetic mean of the z-standardized cognitive ability composites, requiring data to be present for at least two ages (N = 3,939; including 1,261 DZ twin pairs).

Personality and motivation measures. We included 28 self-report measures collected at age 16 (M = 16.48 years; SD = 0.27 years) via self-reports using paper booklet (b) and web-based (w) assessment.

(w) PISA math self-efficacy – eight items (PISA, OECD Programme for International Student Assessment; www.pisa.oecd.org). This scale was selected from the PISA 2000, 2003, and 2006 student questionnaires, comprising eight items asking participants to rate how confident they feel about having to do mathematical tasks on a 4-point scale from *not at all confident* to *very confident*. For example, solving an equation like: $2(x + 3) = (x + 3)(x - 3)$. The total score was created by taking the mean of the eight items, requiring at least four to be present. The scale has an average reliability of 0.83 across OECD countries (Adams & Wu, 2003). We find similar reliability estimates in the present sample (α = 0.90).

(w) PISA math interest – three items (PISA, OECD Programme for International Student Assessment; www.pisa.oecd.org). This scale was selected from the PISA 2000, 2003, and 2006 student questionnaires. The scale asked participants to rate how interested they were in mathematics on a 4-point scale from *strongly disagree* to *strongly agree*. For example, rating statements such as: "I look forward to my mathematics lessons." The total score was created by taking the mean of the three items, requiring at least two to be present. The mean reliability across OECD countries is .75 for this measure (Adams & Wu, 2003). We find a slightly better reliability estimate in the present study than that previously reported (α = 0.93).

(w) PISA time spent on math – three items (PISA, OECD Programme for International Student Assessment; www.pisa.oecd.org). This scale was selected from the PISA 2000, 2003, and 2006 student questionnaires. The scale asked participants to rate how much time they typically spent per week studying mathematics from *no time* to *6 hr or more*. For example "Regular lessons in mathematics at my school." The total score was created by taking the mean of the three items, requiring at least two to be present. The mean reliability across OECD countries is 0.76 for this measure (Adams & Wu, 2003). We find slightly lower reliability estimates (α = 0.53) in the current sample.

(w) Academic self-concept – 11 items (Burden, 1998). This scale aims to assess children's perceptions of themselves as learners and problem solvers by asking children to rate themselves on a 5-point scale from *very much like me* to *not at all like me* to statements such as "I know the meaning of lots of words." The total score was created by taking the mean of the 11 items, requiring at least five to be present. The mean reliability across OECD countries is 0.79 for this measure (Adams & Wu, 2003). We find similar reliability estimates (α = 0.84) in the current sample.

(w) Total attitude toward key subjects – three items (PISA, OECD Programme for International Student Assessment; www.pisa.oecd.org). This scale was selected from the PISA 2000, 2003, and 2006 student questionnaires. Participants were asked to answer the question "In general, how important do you think it is for you to do well in the subjects below?" on a 4-point scale from *not at all important* to *very important* for the subjects English, mathematics, and science. The total score was created by taking the mean of the three items, requiring at least two to be present. The mean reliability across OECD countries is 0.79 for this measure (Adams & Wu, 2003). We find lower reliability in our sample (α = 0.45).

(w) School engagement – 19 items (Appleton, Christenson, Kim, & Reschly, 2006). This scale aims to assess children's engagement with the school environment, including teacher-student relations, control and relevance of schoolwork, peer support, and family support for learning. Participants were required to answer questions such as "I enjoy talking to the teachers at my school" and "Students at my school respect what I have to say" on a 4-point scale from *strongly disagree* to *strongly agree*. The total score was created by taking the mean of the 19 items, requiring at least 10 to be present. The reliability of factors in this measure range from 0.76 to 0.88 (Appleton et al., 2006). We find high reliability (α = 0.99) in the current sample.

(w) Big Five Personality (Extraversion, Openness, Agreeableness, Conscientiousness, neuroticism) – 30 items (Mullins-Sweatt, Jamerson, Samuel, Olson, & Widiger, 2006). We used the subscales from this measure, tapping into Extraversion, Openness, Agreeableness, Conscientiousness, and Neuroticism.

Extraversion – six items. Participants were asked to rate where they were on a scale that varied for each item. For example for the trait *Activity* they had to rate where they were on a scale from *vigorous, energetic, active* to *passive, lethargic.* The total score was created by taking the mean of the five items, requiring at least three to be present. Across five studies, the reliability of this dimension has been estimated to be between 0.60–0.76. In the

current sample, the reliability is within the range of previous studies ($\alpha = 0.68$).

Openness – six items. Participants were asked to rate where they were on a scale that varied for each item. For example for the trait *Fantasy* they had to rate where they were on a scale from *dreamer, unrealistic, imaginative* to *practical, concrete*. The total score was created by taking the mean of the five items, requiring at least three to be present. Across five studies, the reliability of this dimension ranged between 0.51–0.69. In the current sample, the reliability is within the range of previous studies ($\alpha = 0.61$).

Agreeableness – six items. Participants were asked to rate where they were on a scale that varied for each item. For example for the trait *Compliance* they had to rate where they were on a scale from *docile, cooperative* to *oppositional, combative, aggressive*. The total score was created by taking the mean of the five items, requiring at least three to be present. Across five studies, the reliability of this dimension ranged between 0.56–0.72. In the current sample, the reliability is within the range of previous studies ($\alpha = 0.65$).

Conscientiousness – six items. Participants were asked to rate where they were on a scale that varied for each item. For example for the trait *Self-discipline* they had to rate where they were on a scale from *dogged, devoted* to *hedonistic, negligent*. The total score was created by taking the mean of the five items, requiring at least three to be present. Across five studies, the reliability of this dimension ranged between 0.73–0.78. In the current sample, the reliability is within the range of previous studies ($\alpha = 0.77$).

Neuroticism – six items. Participants were asked to rate where they where on a scale that varied for each item. For example for the trait *Angry hostility* they had to rate where they were on a scale from *angry, bitter* to *even-tempered*. The total score was created by taking the mean of the five items, requiring at least three to be present. Across five studies, the reliability of this dimension ranged between 0.62–0.69. The reliability is in line with previous estimates ($\alpha = 0.70$).

(w) Ambition – five items (Duckworth & Quinn, 2009). This measure required participants to rate statements such as "I aim to be the best in the world at what I do" and "I am ambitious" on a 5-point scale from *very much like me* to *Not like me at all*. The total score was created by taking the mean of the five items, requiring at least three to be present. The questionnaire from which these questions were drawn has good reliability, with Cronbach's alphas ranging from 0.83–0.84 (Duckworth & Quinn, 2009). The reliability in the present sample is slightly lower than estimates from previous studies, but is still considered acceptable ($\alpha = 0.74$).

(w) Grit – nine items (Duckworth & Quinn, 2009). This measure required participants to rate statements such as "I am driven to succeed" on a 5-point scale from *very much like me* to *not like me at all*. The total score was created by taking the mean of the nine items, requiring at least five to be present. The questionnaire has good reliability, with Cronbach's alphas ranging from 0.83–0.84 (Duckworth & Quinn, 2009). The reliability in the present sample is slightly lower than estimates from previous studies, but is still considered acceptable ($\alpha = 0.74$).

(w) Curiosity - seven items (Kashdan, Rose, & Fincham, 2004). This measure required participants to rate statements such as "everywhere I go, I am looking out for new things or experiences" and "I would describe myself as someone who actively seeks as much information as I can in a new situation" on a 7-point scale from *strongly agree* to *strongly disagree*. The total score was created by taking the mean of the seven items, requiring at least four to be present. Across five studies, the Cronbach's alpha ranged from 0.72 – 0.80 (Kashdan et al., 2004). In the current sample, the reliability is within the range of previous studies ($\alpha = 0.74$).

(w) Hopefulness – six items (Snyder et al., 1997). This measure required participants to rate sentences about themselves, such as "I think I am doing pretty well" and "I think the things I have done in the past will help me in the future" from *all of the time* to *none of the time*. The total score was created by taking the mean of the six items, requiring at least three to be present. Across eight studies, Cronbach's alpha ranged from 0.72 to 0.86, with a median alpha of 0.77 (Snyder et al., 1997). In the current sample, the reliability is within the range of previous studies ($\alpha = 0.83$).

(b) Strengths and Difficulties Questionnaire: Behavior problems – 20 items (Goodman, 1997). This is a dimensional and developmental measure of child mental health for children aged 3–16 years. Children are required to answer statements on a 3-point Likert scale (*not true*; *quite true*; *very true*). It taps into four domains, each of which are measured by five items, requiring at least three to be present from the subscale:

Conduct problems. For example: 'I get very angry and often lose my temper.' Reliability estimates across studies range from 0.44–0.62 (Mieloo et al., 2012). We found reliability estimates in line with those from other studies ($\alpha = 0.53$).

Hyperactivity/inattention. For example, "I am easily distracted, I find it difficult to concentrate." Reliability estimates across studies range from 0.75–0.87 (Mieloo et al., 2012). Our reliability estimate was in line with those reported in previous studies ($\alpha = 0.73$).

Peer relations. For example, "I have one good friend or more." Reliability estimates across studies range from 0.40–0.58 (Mieloo et al., 2012). In the current sample, the reliability is within the range of previous studies ($\alpha = 0.56$).

Prosocial behavior. For example, "I try to be nice to other people. I care about their feelings." Reliability estimates across studies range from 0.59–0.82 (Mieloo et al., 2012). In the current sample, the reliability is within the range of previous studies ($\alpha = 0.67$).

(b) Strengths and Weaknesses of ADHD Symptoms and Normal Behavior Scale – 18 items (Swanson et al., 2012). This behavior rating scale is based on *Diagnostic and Statistical Manual of Mental Disorders*, fifth edition (*DSM–5*) criteria for ADHD diagnosis measuring inattentive, hyperactive, and impulsive behaviors. Children are asked to compare themselves with other people of their age on a 7-point scale from *far below average* to *far above average*:

Inattention scale. Derived from nine items, an example item is "I sustain attention on tasks or leisure activities," requiring at least half of the items to be present. This scale is scored so that higher scores mean better attention. The reliability for this subscale is 0.91 in one English study and 0.92 in a Spanish study, with good test–retest reliability as well ($r = 0.72$ and 0.49; Lakes, Swanson, & Riggs, 2012). Our reliability estimate was in line with those reported in previous studies ($\alpha = 0.88$).

Hyperactivity scale. Derived from nine items, an example item is "I sit still (control movement of hands/feet)," requiring at

least half of the items to be present. This scale is scored so that higher scores indicate calm and controlled behavior. The reliability for this subscale is 0.93 in one English study and 0.95 in a Spanish study, with good test–retest reliability ($r = 0.71$ and 0.61; Lakes et al., 2012). Our reliability estimate was in line with those reported in previous studies ($\alpha = 0.90$).

(w) Gratitude - six items (McCullough, Emmons, & Tsang, 2002). This measure required participants to rate statements such as "I am grateful to a wide variety of people" and "I have so much in life to be thankful for" on a 7-point scale from *strongly agree* to *strongly disagree*. The total score was created by taking the mean of the six items, requiring at least three to be present. The internal consistency reliability of this scale is 0.82 (McCullough et al., 2002). The reliability is slightly lower than estimates from previous studies, but is still considered acceptable ($\alpha = 0.75$).

(b) Cognitive disorganization – 11 items (Mason, Linney, & Claridge, 2005). This scale, measuring poor attention and concentration, requires individuals to answer 11 items by answering either *yes* or *no*. For example: "Do you frequently have difficulty in starting to do things?"; "Do you find it difficult to keep interested in the same thing for a long time?"; "Is it hard for you to make decisions?" A total score is derived by taking the mean of the 11 items, requiring at least six items to be nonmissing. Reliability of this scale is good, with Cronbach's alpha estimates of 0.77 (Mason et al., 2005). We found the reliability of this scale to be the same as reported previously ($\alpha = 0.77$).

(b) Childhood Anxiety Sensitivity Index – 18 items (Silverman, Fleisig, Rabian, & Peterson, 1991). This is a child-reported questionnaire measuring anxiety sensitivity (i.e., the belief that anxiety symptoms have negative consequences). Responses are rated on a 3-point Likert scale (*not true; quite true; very true*). For example: "I do not want other people to know when I feel afraid"; "I get scared when I feel nervous." A total score is derived by taking the mean of the 18 items, requiring at least nine items to be nonmissing. Reliability of this scale has been tested in clinical and nonclinical samples, both showing good Cronbach's alphas of 0.87 (Silverman et al., 1991). We found the reliability of this scale to be very similar to previous reports of reliability ($\alpha = 0.86$).

(b) Moods and Feelings Questionnaire (MFQ) Short version – 11 items (Angold, Costello, Messer, & Pickles, 1995). A brief questionnaire based on *DSM–III–R* criteria for depression. It is measured on a 3-point Likert scale (*not true*; *quite true*; *very true*) and includes a series of descriptive phrases regarding how the participant has been feeling or acting recently. For example: "I felt I was no good anymore"; "I felt lonely"; "I hated myself." A total score is derived by taking the mean of the 11 items, requiring at least six items to be nonmissing. This scale was reversed so that higher scores meant participants felt fewer depressive traits. The reliability of this scale is good, for both the child version ($\alpha = 0.85$) and the adult version ($\alpha = 0.87$; Angold et al., 1995). We found the reliability of this scale to be in line with previous reports of reliability of this scale ($\alpha = 0.86$).

(w) Life satisfaction – 21 items (Huebner, 1994). This measure taps into different elements of life satisfaction, such as family, school, environment, and life satisfaction from friends. It is measured on a 6-point scale from *strongly agree* to *strongly disagree* and asks participants to rate statements such as: "I enjoy being at home with my family" and "I like where I live." A total score is derived by taking the mean of the 21 items, requiring at least 11 items to be nonmissing. Previous studies have shown the reliability of this measure to be good, estimated at $\alpha = 0.92$ (Huebner, 1994). In the present sample, we found a similar estimate ($\alpha = 0.86$).

(w) Subjective happiness – four items (Lyubomirsky & Lepper, 1999). These questions tap into perceived happiness, asking participants to complete a sentence. For example: "In general, I consider myself . . ." with a 7-point response option from . . . *not a very happy person* to . . . *a very happy person*. A total score is derived by taking the mean of the four items, requiring at least two items to be nonmissing. Reliability estimates from 14 samples ranged from 0.79–0.94 (Lyubomirsky & Lepper, 1999). We found the reliability of this scale in our sample to be similar to previously reported estimates ($\alpha = 0.78$).

(w) Optimism – six items (Scheier, Carver, & Bridges, 1994). This measure required participants to rate statements such as "In uncertain times, I usually expect the best" and "I'm always optimistic about my future" on a 5-point scale from *very much like me* to *not like me at all*. The total score was created by taking the mean of the six items, requiring at least three to be present. The reliability of this measure is good, estimated at $\alpha = 0.82$ (Scheier et al., 1994). We found the reliability of this scale in our sample to be similar to previously reported estimates ($\alpha = 0.76$).

Supplemental Table S1 shows that for most measures, there were small but significant gender differences, and that for some measures there were small effects of age. Prior to any further analyses, all variables were corrected for the effects of gender and age using the regression method to obtain *z*-standardized residuals.

Because of the large number of measures and the widespread correlations (Supplemental Figure S1), we looked at empirical studies of personality structure and conducted factor analysis (FA) to reduce the large number of measures to six domains. These comprised: Extraversion, Neuroticism, Openness to Experience, Conscientiousness, Agreeableness, and Academic Motivation.

Before conducting factor analysis, we performed parallel analysis to guide factor extraction. In parallel analysis, FA is repeatedly applied to sets of randomly generated, uncorrelated data. These data contain the same sample parameters as in the study sample, and by simulating numerous FAs parallel analysis produces a distribution of eigenvalues. If the component eigenvalue in the study sample is greater than the 95th percentile of the simulated eigenvalues, the retention of this component is justified (O'Connor, 2000). Results from parallel analysis based on our sample parameters ($N = 603$, based on the total number of individuals with no missing data; number of variables = 28; number of iterations = 1,000) indicated the retention of five factors (see online supplementary Figure S2). To guide our decision-making in creating personality domains, we performed oblique rotation (promax) to allow for correlated factors.

The five-factor FA accounted for 42% of the total variance. Factor loadings revealed an underlying structure representing the FFM. However, instead of an Extraversion factor, there was a factor representing Academic Motivation. The measure of extraversion loaded substantially onto factors of Openness (0.59), Neuroticism (−0.25), and Conscientiousness (−0.26) instead of forming a separate factor. This is presumably because there were no other scales that served as indicators of Extraversion. Based on existing scientific knowledge of personality structure, we decided to rerun the FA excluding Extraversion and instead have Extraversion as its own separate personality

Table 1
Factor Analysis

Extracted factors	Rotation sums of squared loadings		
	Total	% of variance	Cumulative %
1	3.085	11.426	11.426
2	2.824	10.459	21.885
3	2.288	8.472	30.358
4	1.586	5.875	36.233
5	1.566	5.799	42.032

Note. Individual and cumulative variance explained by the top five factors.

domain. Repeated parallel analysis confirmed the selection of the top five factors for rotation. The final FA without extraversion also explained 42% of the total variance (see Table 1), and item loadings revealed five factors: Neuroticism (e.g., cognitive disorganization and anxiety), Openness to Experience (e.g., ambition and curiosity), Conscientiousness (e.g., attention and focus), Agreeableness (e.g., prosocial behavior and gratitude), and Academic Motivation (e.g., math self-efficacy and engagement with key subjects). Item loadings are shown in Table 2.

Rather than extracting factor loadings to create personality domains for subsequent analysis, which would lead to a substantial loss of data due to listwise deletion, we created variables by taking the arithmetic mean of the standardized subscales, requiring at least half to be present and reversing measures when they correlated negatively with a factor. Composites based on factor loading extraction and mean composite calculation correlated highly (average $r = .91$). Descriptive statistics of the six personality and motivation domains and the 28 subscales are shown in Supplemental Table S1, and correlations between the domains can be found in Supplemental Figure S3.

To test whether there were any meaningful differences between those with missing and nonmissing personality and motivation composites, we conducted sensitivity analyses. We assessed mean differences in socioeconomic status assessed at first contact (mean composite of parental education, occupation, and maternal age at the birth of the first child), general cognitive ability and GCSE results between missing and nonmissing personality and motivation composites scores. We found small differences between those with missing and nonmissing data, accounting for an average of 1% (range 0.1% to 2.6%) of the phenotypic variance (see Supplemental Table S2).

Statistical Analyses

Genome-wide polygenic score calculation. For the 10,346 individuals in our sample, we calculated three polygenic scores. The first was based on the summary statistics for a GWA meta-analysis for years of education ($N = 766{,}345$ after removal of all 23andme participants; Lee et al., 2018). The second and third were based on the two largest GWA meta-analyses for personality traits to date, Neuroticism ($N = 329{,}821$; Luciano et al., 2018) and Well-being ($N = 298{,}420$; Okbay, Baselmans, et al., 2016).

Table 2
Rotated Item Factor Loadings

Measure	Neuroticism	Openness	Conscientiousness	Agreeableness	Academic Motivation
Cognitive disorganization	**0.52**		0.49		
Anxiety	**0.62**				
MFQ	**0.80**				
Subjective happiness	**−0.67**			0.21	
Life satisfaction	**−0.68**			0.34	
Peer problems	**0.54**				
Academic self-concept		**0.50**			0.36
Ambition		**0.68**			
Curiosity		**0.71**			
Hopefulness	−0.25	**0.72**			
SDQ hyperactivity			**0.80**		
SWAN hyperactivity			**−0.67**		
SWAN inattention		0.21	**−0.61**		
Agreeableness				**0.68**	
SDQ prosocial				**0.51**	
Maths self-efficacy					**0.96**
Maths interest					**0.54**
Maths time spent					**0.25**
Attitudes key subjects		0.20			**0.18**
School engagement				**0.21**	
Openness		**0.24**	0.22		
Conscientiousness	0.22	0.35	**−0.39**		
SDQ conduct	0.20	0.34	0.27	**−0.48**	
GRIT		**0.44**	−0.41		
Gratitude	−0.21	0.26		**0.43**	
Neuroticism	**0.45**	−0.31			
Optimism	−0.29	**0.46**			

Note. MFQ = Moods and feelings questionnaire; SDQ = Strengths and difficulties questionnaire; SWAN = Strengths and weaknesses of ADHD symptoms and normal behavior rating scale. Oblique (promax) rotation was applied. Only variables with factor loadings of ≥0.20 are shown. Measures included in the same composite are set in bold.

The first wave of TEDS genotyped samples (N = 2,148; Trzaskowski et al., 2013) was included in the discovery sample of the Well-being GWA meta-analysis. Therefore, we performed a statistical correction on the summary statistic effect size coefficients and p values (Socrates et al., 2017) to account for the overlap between the discovery and target sample. We first replicated the genome-wide association study on Well-being using genotypes from the 2,148 TEDS individuals that were included in the meta-analysis, following the GWA protocol applied in the discovery analysis (Okbay, Baselmans, et al., 2016). Second, the obtained beta coefficients and standard errors for each SNP were then used to adjust the meta-analysis beta coefficients and standard errors. These adjusted values are analogous to the effects for each SNP if the TEDS sample would have been removed in the discovery meta-analysis (Socrates et al., 2017). Third, we calculated new p values based on the adjusted beta coefficients and standard errors. The adjusted summary statistics for well-being were used for polygenic score calculation in the full TEDS sample.

A GPS is calculated by using information from GWA study summary statistics about the strength of association between a genetic variant and a trait, to score individuals' genotypes in independent target samples such as TEDS. Here, we used a Bayesian approach to polygenic score calculation, implemented in the software *LDpred* (Vilhjálmsson et al., 2015). In comparison with conventional clumping and p value thresholding approaches, *LDpred* has demonstrated an improvement in predictive accuracy (Vilhjálmsson et al., 2015). Through this method, a posterior effect size is calculated for each single SNP that is present in both the GWA study summary statistics and the target genotype sample. To calculate this, the original summary statistic effect size estimates are adjusted based on two factors: (a) the relative influence of a SNP given its level of LD with surrounding SNPs in the target sample (here TEDS), and (b) a prior on the effect size of each SNP. This prior depends on the SNP-heritability of the discovery (i.e., GWA study) trait and an assumption on the fraction of causal markers believed to influence the discovery trait. For this study, we set the LD radius to a 2 Megabase window and used a prior based on a fraction of causal markers of 1, meaning that we apply the assumption that all SNPs are causally influencing the discovery trait. Therefore, the prior reweights the beta effect sizes such that the effects are spread out among the SNPs across the whole genome in proportion to the LD present among these SNPs. To accommodate the high computational demands of these calculations, we reduced our genotype data set to SNPs that had perfect imputation scores (info = 1), leaving 515,100 SNPs for analysis.

In the next step, all trait-associated alleles were counted (0, 1, or 2 for each SNP), weighted by the posterior SNP effect size obtained through *LDpred*, and summed across the genome to calculate a GPS for each individual in TEDS. Although we use a prior based on a fraction of causal markers of 1 to create a GPS for the main analysis, we calculated two more scores with fractions 0.01 and 0.10 for comparison (see online supplementary figures S4–S6 for correlations based on different priors).

To control for platform effects (Affymetrix vs. Illumina) and plate effects, as well as effects of population stratification, we regressed all GPS used in this study on platform and plate data, and the first 10 principal components. For all subsequent analyses, we used z-standardized residuals.

Trait prediction based on regression analysis. To test the extent to which *EduYears* GPS, neuroticism GPS, and well-being GPS can predict personality traits that are related to GCSE, we applied regression analysis, using robust standard errors to account for the clustering in our data. Because these traits are associated with general cognitive ability, we repeated these analyses using the residuals obtained from regressing our personality and motivation traits on general cognitive ability. We performed bootstrapping with 10,000 bootstrap samples to obtain 95% bootstrap percentile intervals for each coefficient of determination (R^2). To identify whether prediction estimates between the three GPS differed significantly, we used the Williams modification of the Hotelling test (Williams, 1959), which takes into account nonindependence of the predictor variables. Additionally, we performed three multiple regression analyses with the polygenic scores as outcomes to assess the relative contributions of general cognitive ability and the personality and motivation phenotypes to polygenic score variation.

Sensitivity analyses for GPS trait prediction. We carried out two types of sensitivity analyses. First, by virtue of the considerable GWA study sample sizes differences between EduYears ($N \sim$ 760,000) and the personality association studies (neuroticism: $N \sim$ 330,000; well-being; $N \sim$ 300,000), it is possible that differences in GPS predictions are a product of differences in power to detect effect sizes. We therefore repeated our association analyses between *EduYears* GPS and personality measures using the 2016 GWA study summary statistics based on a sample of ~300,000 individuals to assess any gains in prediction as a result of the steep sample size increase (Supplemental Figure S7).

Second, it is a common concern that regression coefficients from GPS analyses are biased due to overfit to the data (Choi, Mak, & O'Reilly, 2018; Wray et al., 2013). Because of the lack of an independent validation sample to test model performance, we carried out internal validation by applying repeated fivefold cross-validation in our sample to reduce model bias and variability of cross-validation prediction estimates (Kim, 2009). Furthermore, we restricted our sample to unrelated individuals only to simultaneously assess a potential bias due to the inclusion of relatives in our target sample (for descriptive statistics of the unrelated sample, see Supplemental Table S3). For each of the folds, the sample was randomly partitioned into 80% training samples, used to train the model, and 20% validation samples used to evaluate the model performance, where each individual appeared only once in the validation sample. The fivefold cross-validation procedure was repeated 50 times with random data splits, and the final cross-validated R^2 estimates were calculated as the average of all model estimates.

GPS prediction of covariance. Finally, we calculated the extent to which each GPS accounts for the relation between personality and motivation domains and GCSE grades using structural equation modeling. We estimated (i) Effect of the GPS on the personality/motivation traits and GCSE grades (a^*b), (ii) the residual correlation between personality/motivation traits and GCSE results after accounting for the mutual effect of the GPS on both traits (c'), and (iii), the total covariance explained by the model ($a^*b + c'$). Using this information, it is possible to calculate the extent to which a GPS explains the association between personality/motivation domains and GCSE results (a^*b) / ($a^*b + c'$); see Supplemental Methods S1).

Alpha correction for multiple testing. Multiple testing was accounted for by adjusting the significance threshold by the effective number of tests in accordance with the Nyholt-Šidák correction, which accounts for correlation among the variables. For the Nyholt approach, eigenvalue decomposition is applied to a correlation matrix containing the variables used for analysis, and the eigenvalue variance in relation to the absolute number of variables is used to calculate the effective number of variables (D_{eff}; Nyholt, 2004). For our analyses, we calculated an effective number of variables based on seven input variables (GCSE results and six personality variables) before and after correcting these variables for general cognitive ability, resulting in D_{eff} of 6.27 and 6.34, respectively. These derived values were then used to calculate the Šidák corrected (Šidák, 1971) significance threshold ($\alpha = 1 - 0.95^{1/Deff}$). We calculated a total number of 58.83 tests performed for our main analyses. This was calculated by adding together the number of tests: 18.81 tests for comparing each of the three GPS with the seven variables (3×6.27), 19.02 tests for comparing the three GPS with the seven variables while accounting for general cognitive ability (3×6.34), 18 tests to calculate the extent to which the three GPS account for the covariance between GCSE grades and personality traits (3×6) and three multiple regressions (3). This resulted in a corrected p value threshold of 8.72×10^{-4}.

All analyses were performed using the statistical software R (R Core Team, 2017). Parallel analysis was performed using the 'parallel' function in the package *nFactors (*Raiche & Magis, 2010). Factor analysis was performed using the 'factanal' function in the *stats* package. Bootstrapping was performed using the 'boot' function in the *boot* package (Canty & Ripley, 2012). Robust standard errors were calculated using the 'coeftest' function implemented in the *lmtest* package (Zeileis & Hothorn, 2002). Significance of difference between correlation coefficients was tested using the 'r.test' function in the *psych* package (Revelle, 2017). Repeated cross-validation was performed using the 'trainControl' and 'train' function (method 'lm') in the package *caret* (Kuhn, 2015). Structural equation modeling analyses were performed using the package *lavaan* (Rosseel et al., 2018), selecting the robust standard error option to account for the clustering in our data due to the inclusions of DZ twin pairs.

Results

Correlations Between Personality Domains and Academic Achievement

Phenotypic correlations between academic achievement (GCSE results) and the six personality and motivation domains were examined to evaluate the strength of associations between these measures. Pearson's correlation coefficients were statistically significant and absolute values ranged from 0.13 to 0.45 (see Supplemental Figure S3). For correlations between all underlying personality facets and motivation traits and GCSE results, see Supplemental Figure S3.

Polygenic Score Prediction of Personality and Academic Motivation

To test the predictive validity of the polygenic score for years of education (*EduYears* GPS) and the six personality and motivation domains that contribute to educational success, we performed association analyses. Figure 1A shows that *EduYears* GPS was a significant predictor of all personality/motivation domains but Neuroticism and Extraversion, which did not withstand correction for multiple testing. *EduYears* GPS was significantly positively associated with Agreeableness ($\beta = 0.098$, $p = 2.17 \times 10^{-16}$, $R^2 = 0.010$), Conscientiousness ($\beta = 0.077$, $p = 5.59 \times 10^{-5}$, $R^2 = 0.006$), Openness ($\beta = 0.141$, $p = 5.09 \times 10^{-16}$, $R^2 = 0.021$), and Academic Motivation ($\beta = 0.167$, $p = 3.99 \times 10^{-21}$, $R^2 = 0.029$). The direction of associations indicated that higher *EduYears* GPS scores related to higher Academic Motivation, Openness, Conscientiousness and Agreeableness. We also tested the association with GCSE grades, finding EduYears GPS significantly predicted GCSE results ($\beta = 0.370$, $p = 3.36 \times 10^{-288}$, $R^2 = 0.137$), as reported in Allegrini et al., 2018.

The GPS for neuroticism significantly negatively related to GCSE results ($\beta = -0.067$, $p = 1.51 \times 10^{-9}$, $R^2 = 0.004$), Openness ($\beta = -0.065$, $p = 4.37 \times 10^{-3}$, $R^2 = 0.004$), and Academic Motivation composites ($\beta = -0.088$, $p = 6.43 \times 10^{-7}$, $R^2 = 0.007$), and was as expected positively associated with the Neuroticism composite ($\beta = 0.087$, $p = 2.21 \times 10^{-11}$, $R^2 = 0.007$; Figure 1A). Associations with the Conscientiousness, Extraversion, and Agreeableness composite did not survive multiple testing corrections. Overall, the direction of effects indicated that individuals that carry more genetic variants that are related to Neuroticism (i.e., individuals with a higher Neuroticism GPS) scored higher on Neuroticism, had significantly lower GCSE grades, and showed a significant decrease in Openness and Academic Motivation.

The well-being GPS was a significant predictor of the Neuroticism composite ($\beta = -0.076$, $p = 1.74 \times 10^{-8}$, $R^2 = 0.006$) and the Agreeableness composite ($\beta = 0.053$, $p = 2.97 \times 10^{-5}$, $R^2 = 0.003$), such that a higher well-being GPS related to lower Neuroticism scores, and higher Agreeableness scores. No correlation was found with GCSE score (Figure 1A). Results for other GPS thresholds are reported in Supplemental Figures S4-6.

With the exception of the Neuroticism composite and Extraversion, the magnitudes of the correlation coefficients between *EduYears* GPS and the personality measures were at least twice as high as of those relating to the Neuroticism and Well-being GPS. Formal comparisons between correlation coefficients showed that *EduYears* GPS was a significantly stronger predictor than the Neuroticism and Well-being GPS for GCSE results ($p = 1.00 \times 10^{-109}$; $p = 1.90 \times 10^{-138}$, respectively), Openness ($p = 8.8 \times 10^{-4}$; $p = 3.00 \times 10^{-6}$, respectively), and Academic Motivation ($p = 3.80 \times 10^{-4}$; $p = 1.40 \times 10^{-10}$, respectively). For Agreeableness, *EduYears* GPS was a better predictor than the Neuroticism GPS ($p = 2.30 \times 10^{-6}$), but not the Well-being GPS ($p = .006$). The contrasts between the Neuroticism and the Well-being GPS showed that the Neuroticism GPS significantly predicted more variance in Academic Motivation ($p = 7.90 \times 10^{-4}$) and GCSE results ($p = 3.20 \times 10^{-5}$).

Controlling for General Cognitive Ability

General cognitive ability correlated with personality and motivation facets and composites, as well as GCSE grades (Supplemental Figure S3). Therefore, we corrected the composites and GCSE results for variance explained by general cognitive ability

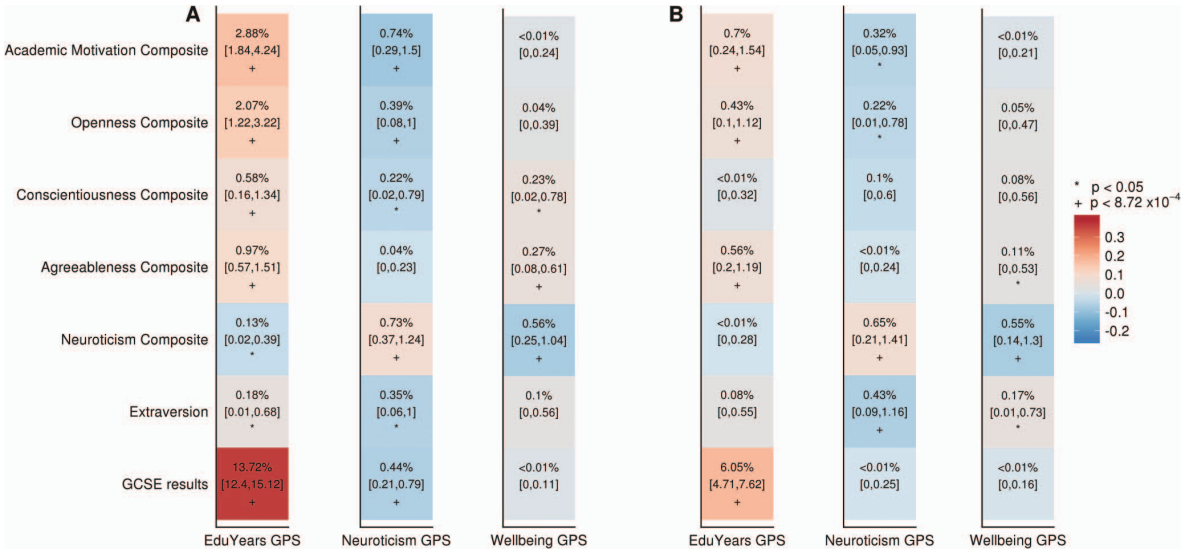


Figure 1. Genome-wide polygenic scores (GPS) as a predictor of six personality and motivation domains and GCSE (General Certificate of Secondary Education) results (A) before and (B) after accounting for general cognitive ability. The color shading represents magnitude of Pearson Correlation coefficients, and the values in each cell represent the amount of phenotypic variance explained by the polygenic scores. Values in square brackets represent the lower and upper bounds of the 95% bootstrapped percentile intervals based on 10,000 bootstrap samples. '+' = p value threshold for significance after correction for multiple testing (8.72×10^{-4}). See the online article for the color version of this figure.

and repeated the association analyses as shown in Figure 1B. We found that *EduYears* GPS was still a significant, albeit attenuated, predictor of GCSE grades, Agreeableness, Openness, and Academic Motivation. For the Neuroticism GPS, previously significant correlations with Academic Motivation and Openness did not reach the multiple-testing corrected p value threshold after accounting for general cognitive ability, and the magnitudes of associations were mostly attenuated for GCSE results. In contrast, the associations with Extraversion and Neuroticism remained significant and of similar strength after correction for general cognitive ability. The correlation between the Well-being GPS and the Neuroticism composite remained statistically significant, with no attenuation of effect size. These results suggest that the covariance shared between the GPS and the personality and motivation domains is partly tagged by general cognitive ability, but not solely explained by it. Attenuations were substantially more pronounced for *EduYears* GPS associations (71.3% including GCSE; 73.9% excluding GCSE) than for the neuroticism (50.9% including GCSE; 43.2% excluding GCSE) and Well-being GPS (4.5% including GCSE; 5.2% excluding GCSE), indicating that, as expected, the *EduYears* GPS tags more genetic variants related to general cognitive ability.

Sensitivity Analyses

Associations between the 2016 *EduYears* GPS and personality measures. To assess the extent to which the considerably larger GWA study sample size had on *EduYears* GPS predictions of personality traits relative to the neuroticism and well-being GPS, we repeated our analyses using a GPS that is based on the 2016 *EduYears* GPS that has a similar sample size to the neuroticism and well-being GWA study. We found that for the personality domains, Pearson's correlation coefficients using the 2016 and the 2018 *EduYears* GPS were almost identical (Supplemental Figure S7), indicating that GWA study power differences between *EduYears and* neuroticism and well-being are not likely to explain the differences in predictions of personality measures.

Repeated cross-validation of prediction estimates. To test whether our regression model estimates were biased, potentially because of overfit data or relatedness within the sample, we contrasted them to more robust estimates obtained from repeated fivefold cross-validation in unrelated samples (see Figure 2). Model estimates derived from our previous analyses using the full sample were very similar to the mean of all cross-validated predictions, and without exception fell within the 95% cross-validated R^2 percentile ranges. Moreover, where prediction estimates from our full sample differed, the values were generally more conservative than the mean cross-validated R^2 values. Overall, these comparisons indicate that our model predictions in our full sample are not inflated due to overfitting or relatedness.

Multiple regression analyses predicting polygenic scores from cognitive ability, personality and Academic Motivation. To further assess the contributions of cognitive ability and the personality/motivation domains in the polygenic score variation, we performed multiple regression analyses with the polygenic scores as dependent variables. Table 3 shows the beta coefficients for each measure in the joint prediction models. Results for Model

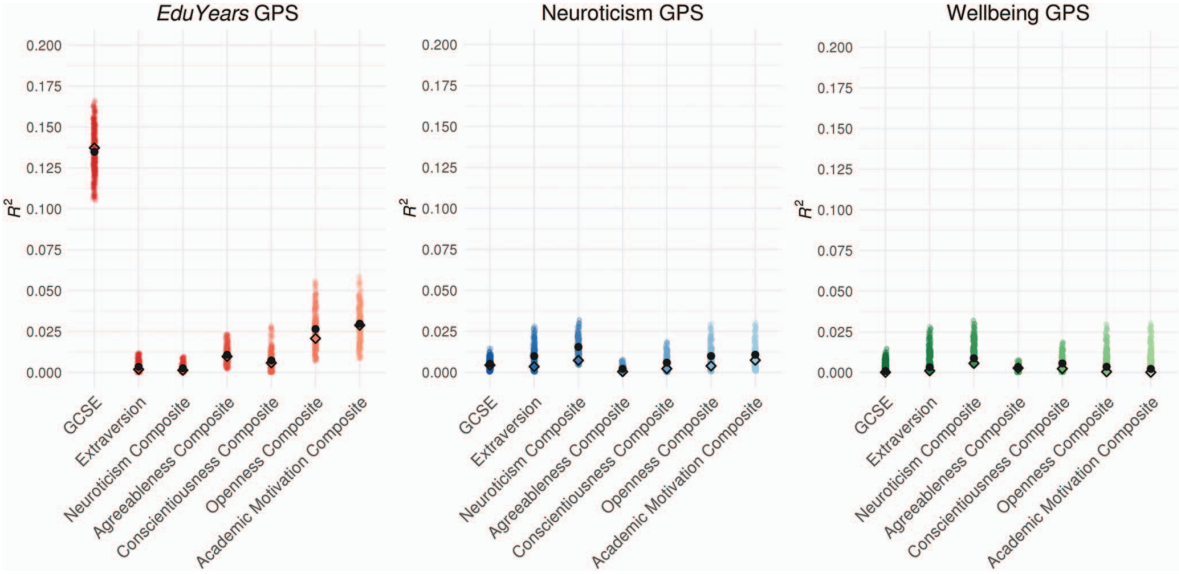


Figure 2. Comparison of prediction in full sample including DZ twin pairs to cross-validated predictions in unrelated individuals. We performed fivefold cross-validation with 50 repetitions. R^2 estimates that fall within the 95% percentile of all cross-validation prediction estimates are represented by the colored dots. The black dots indicate the mean of these the cross-validated R^2, and the diamond shaped symbols indicate the prediction estimates obtained from the original regression models using the full sample. See the online article for the color version of this figure.

1 indicated that a significant proportion of variance in *EduYears* GPS was explained by the predictors ($F(7, 2149) = 29.00$, $p = 1.94 \times 10^{-38}$, $R^2_{adjusted} = 0.083$). The effects were predominantly driven by general cognitive ability and the Agreeableness composite. The overall multiple regression model predicting neuroticism GPS was significant ($F(7, 2149) = 6.29$, $p = 2.49 \times 10^{-7}$, $R^2 = 0.017$), with the largest effect sizes from individual contributors stemming from general cognitive ability and Neuroticism. The multiple regression model predicting the well-being GPS was not statistically significant ($F(7, 2149) = 3.11$, $p = 2.87 \times 10^{-3}$, $R^2 = 0.007$), and most of the variance was, albeit not significantly, accounted for by the Neuroticism composite.

Polygenic score prediction of covariation. Because GCSE grades, *EduYears* GPS and the personality and motivation domains are intercorrelated (Supplemental Figure S1), we tested the extent to which *EduYears* GPS accounted for the association between GCSE grades and the personality and motivation domains. Figure 3 and Table 4 show that *EduYears* GPS accounted for a significant amount of covariation between GCSE grades and Academic Motivation (12.2%, $p = 1.24 \times 10^{-12}$), Openness (14%, $p = 6.06 \times 10^{-11}$), Conscientiousness (7.7%, $p = 8.72 \times 10^{-4}$), and Agreeableness (15.6%, $p = 8.69 \times 10^{-13}$). For comparison, we performed the same analyses using the Neuroticism and Well-being GPS. The Neuroticism GPS only accounted for a significant

Table 3
Results From Multiple Regression Analyses: Cognitive and Personality/Motivation Composites Predicting Genome-Wide Polygenic Scores

Predictor	Model 1: *EduYears* GPS			Model 2: Neuroticism GPS			Model 3: Well-being GPS		
	β	*SE*	*p*	β	*SE*	*p*	β	*SE*	*p*
General cognitive ability	**.252**	**.024**	**$7.54 \times 10^{-25+}$**	**−.079**	**.023**	**$5.56 \times 10^{-4+}$**	.045	.024	.058
Academic Motivation	.071	.026	.007*	−.042	.026	.102	−.045	.027	.093
Openness	.041	.029	.158	−.010	.030	.734	.020	.031	.506
Conscientiousness	−.059	.027	.032	.013	.027	.627	−.012	.028	.668
Agreeableness	**.115**	**.031**	**$1.71 \times 10^{-4+}$**	.066	.031	.038*	−.019	.034	.570
Neuroticism	.041	.027	.120	**.095**	**.028**	**$6.05 \times 10^{-4+}$**	−.096	.030	$1.37 \times 10^{-3*}$
Extraversion	.009	.025	.719	−.030	.024	.213	<.001	.026	.997

Note. Beta coefficients, standard errors, and *p* values are presented for each of the predictors in the regression models.
* $p < .05$. '+' = 8.72×10^{-4} (*p* value threshold for significance after correction for multiple testing). Results that reached significance after correction for multiple testing are set in bold.

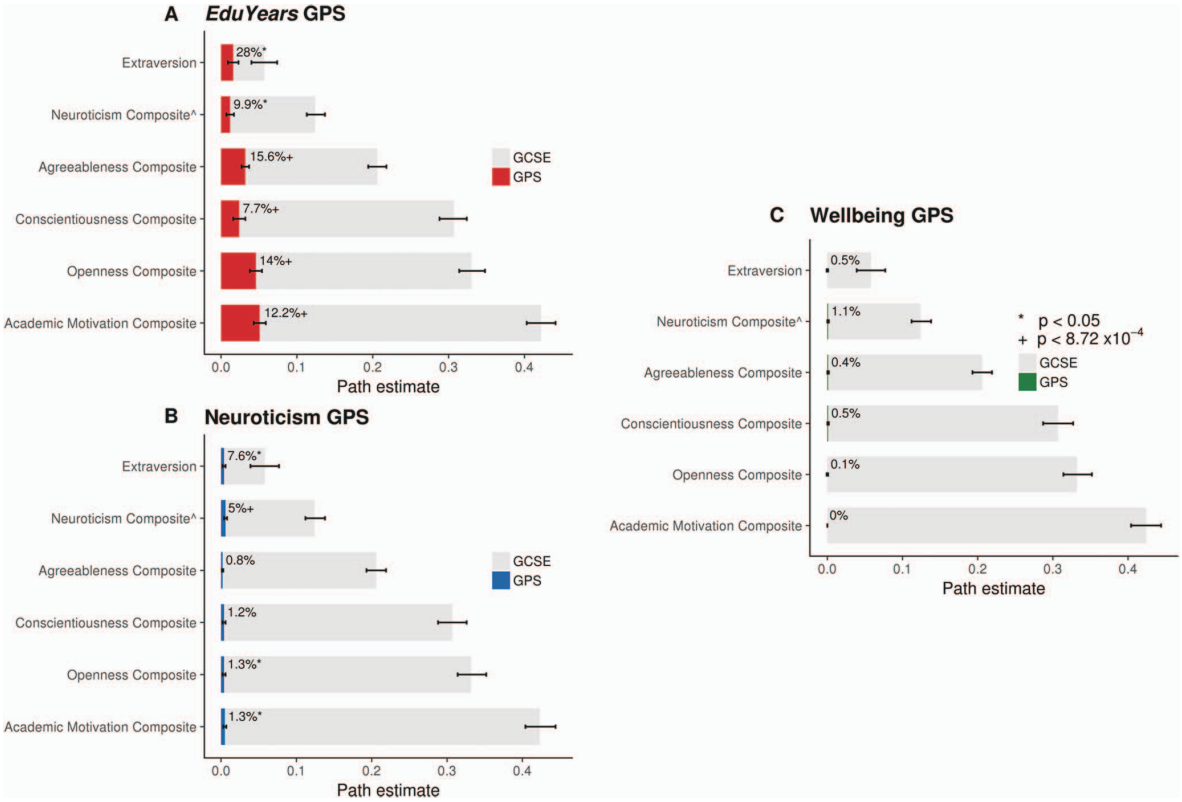


Figure 3. Standardized path estimates for the association between GCSE (General Certificate of Secondary Education) grades and personality/motivation domains, and the proportion of these associations accounted for by *EduYears* GPS (genome-wide polygenic scores) (A) Neuroticism GPS (B), and Well-being GPS (C). Error bars represent robust standard errors. Path estimates presented are estimated based on maximum likelihood (see Table 4 for all path estimates). '+' = p value threshold for significance after correction for multiple testing (8.72×10^{-4}). See the online article for the color version of this figure.

amount of covariance between Neuroticism and GCSE grades (5%, $p = 1.92 \times 10^{-4}$; Figure 3; Table 4). No significant covariance was explained by the Well-being GPS.

Discussion

Summary of Findings

Our results show that a genome-wide polygenic score (GPS) for educational attainment predicts a number of personality and motivation domains, including Agreeableness, Openness, Conscientiousness, and Academic Motivation. We find that the educational attainment GPS (*EduYears*) is more predictive of Academic Motivation, Openness, and Agreeableness than personality GPS themselves, and that *EduYears* GPS explains between 8% and 16% of the covariance between personality and motivation domains and educational achievement at age 16. These findings suggest that DNA variants contributing to educational attainment are also important predictors of personality and motivation.

Much of the previous research using *EduYears* GPS has focused on its relation with cognitive traits, such as general cognitive ability and educational outcomes (Belsky et al., 2018; Lee et al., 2018; Okbay, Beauchamp, et al., 2016; Rietveld et al., 2013; Selzam, Dale, et al., 2017; Selzam, Krapohl, et al., 2017). In contrast, our findings demonstrate the broad, multifaceted nature of *EduYears* GPS, which is also associated with a variety of personality and motivation traits. Indeed, we show that *EduYears* GPS significantly predicts four out of six personality and motivation domains: Academic Motivation, Openness, Conscientiousness, and Agreeableness, explaining between 0.6% and 2.9% of the variance. Our formal comparisons show that for Academic Motivation and Openness, *EduYears* GPS was a better predictor than the neuroticism and well-being GPS, as well as for Agreeableness in comparison to the neuroticism GPS. In predicting Neuroticism and Extraversion, *EduYears* GPS achieves comparable effect sizes to the neuroticism and well-being GPS. Our sensitivity analyses showed that the larger prediction estimates for *EduYears* GPS are not a function of the larger GWA study sample

Table 4
Path Estimates and Standard Errors

Predictor	Parameter	*EduYears* GPS		Neuroticism GPS		Well-being GPS	
		β	Robust *SE*	β	Robust *SE*	β	Robust *SE*
Academic Motivation composite	GPS effect	.051^{+}	.008	.005*	.002	<.001	<.001
	Resid cor	.371^{+}	.019	.418^{+}	.020	.424^{+}	.020
	Total effect	.422^{+}	.020	.424^{+}	.020	.424^{+}	.020
	Proportion	**.122^{+}**	**.017**	.013*	.005	<.001	<.001
Openness composite	GPS effect	.046^{+}	.008	.004*	.002	<.001	<.001
	Resid cor	.284^{+}	.017	.328^{+}	.019	.332^{+}	.019
	Total effect	.331^{+}	.018	.333^{+}	.019	.232^{+}	.019
	Proportion	**.139^{+}**	**.021**	.013	.006	<.001	<.001
Neuroticism composite	GPS effect	−.012*	.005	−.006^{+}	.002	−.001	.001
	Resid cor	−.112^{+}	.012	−.118^{+}	.013	−.123^{+}	.013
	Total effect	−.125^{+}	.013	−.125^{+}	.013	−.123^{+}	.013
	Proportion	.099*	.036	**.050^{+}**	**.013**	.011	.008
Conscientiousness composite	GPS effect	.024*	.009	.004	.002	.001	.001
	Resid cor	.283^{+}	.029	.303^{+}	.019	.306^{+}	.020
	Total effect	.306^{+}	.031	.307^{+}	.020	.306^{+}	.020
	Proportion	**.077^{+}**	**.023**	.013	.007	.004	.004
Agreeableness composite	GPS effect	.032^{+}	.005	.002*	.001	<.001	<.001
	Resid cor	.174^{+}	.012	.204^{+}	.013	.205^{+}	.013
	Total effect	.206^{+}	.013	.206^{+}	.013	.205^{+}	.013
	Proportion	**.156^{+}**	**.022**	.008	.005	.003	.003
Extraversion	GPS effect	.016*	.007	.004*	.002	<.001	<.001
	Resid cor	.041*	.017	.054^{+}	.019	.058^{+}	.019
	Total effect	.057*	.018	.058^{+}	.019	.058^{+}	.019
	Proportion	.278*	.116	.076*	.038	.004	.012

Note. GPS effect = effect of the genome-wide polygenic score (GPS) on both traits; Resid cor = residual correlation between phenotypes after mutually adjusting for the effects of the GPS; Total effect = effect accounted for by the model (resid cor + GPS effect); Proportion = the proportion of the total effect that is accounted for by the GPS effect (GPS effect/total effect). Statistically significant proportions of variance explained are in bold.
* $p < .05$. '+' = 8.72×10^{-4} (p value threshold for significance after correction for multiple testing).

size in comparison with the neuroticism and well-being GWA study, as a GPS for *EduYears* based on the 2016 GWA study with a comparably large sample produced almost identical results.

We find that even once we accounted for general cognitive ability, *EduYears* GPS still predicted significant variance in Agreeableness (0.6%), Openness (0.4%), Academic Motivation (0.7%), and GCSE results (6.1%). Correcting for general cognitive ability substantially attenuated associations between the personality traits and *EduYears* GPS (74%), compared with neuroticism GPS (43%), and even less for the well-being GPS (5%). Attenuation patterns are also mirrored in the multiple regression analyses. We found that general cognitive ability remains a significant predictor for *EduYears* GPS and neuroticism GPS but not the well-being GPS when controlling for all personality measures, and the beta effect sizes are larger for the prediction of *EduYears* than for the neuroticism GPS. One likely explanation for this finding is that the GWA study on years of education tags more general cognitive ability related variants than the neuroticism and well-being GWA study. Therefore, statistically controlling for general cognitive ability in the prediction of personality traits would have a greater impact on *EduYears* GPS compared with either neuroticism or well-being GPS. The finding that *EduYears* GPS is correlated with personality and motivation traits, even after accounting for general cognitive ability is particularly interesting for two reasons. First, it shows that a polygenic score for years of education not only tags genetic variance associated with its target trait, but also many other traits that contribute to how long a person stays in education. And second, it illustrates that staying in education depends on more than just intelligence; many cognitive and noncognitive genetically influenced traits contribute to educational attainment.

In addition to showing that *EduYears* GPS explains significant variance in personality and motivation domains, we also show that it explains between 8% and 16% of the associations between personality and motivation domains and educational achievement at age 16. In contrast, the well-being GPS did not significantly account for any covariance between these traits and GCSE results, and the neuroticism GPS accounted for a significant amount of variance only in Neuroticism (5%). As previously mentioned, a possible explanation for this finding is that GWA studies performed on personality traits may tag variants specific to the target trait, rather than capturing trait-related variants that also contribute to the development of skills important for educational achievement. In contrast, a GWA study performed on educational attainment is likely to capture genetic variants that are important contributors to many down-stream educationally relevant traits. For example, if motivation is a genetically influenced trait and an important factor for higher educational attainment, a GWA study on years of education will indirectly capture some of the genetic effects relating to motivation if individuals with higher motivation levels are likely to stay in education for longer on average. Another possible mechanism to explain these associations may be that passive rGE is more pronounced for educational attainment than for neuroticism and well-being. It has been shown that nontransmitted genetic variants related to educational attainment in parents predict their children's educational achievement, in addition to their children's inherited genetic propensities for educational at-

tainment (Kong et al., 2018). This finding points toward a source of passive rGE, where parents provide a family environment based on their own genetics, which in turn contributes to their children's development, even if they do not share these same genetic markers with their parents. A GWA study on educational attainment might therefore pick up on both the direct effects between the individuals' genetic markers and their educational attainment, and also the effects of the family environment that covaries with their parental nontransmitted genotypes. Therefore, part of the associations we find could be reflecting passive rGE.

Overall, our results demonstrate the substantial genetic pleiotropy (i.e., one DNA marker affects several traits) across educational achievement and educationally relevant traits, although it is not possible to distinguish between biological pleiotropy (i.e., one DNA marker directly affects several traits) and mediated pleiotropy (i.e., one DNA marker directly affects one trait, which then in turn affects another trait) (Solovieff et al., 2013). The findings of this study support previous twin research, showing that between 8% and 37% of the covariance between personality traits and GCSE results is explained by shared genetic factors (Krapohl et al., 2014). Although the difference between the magnitudes of effect sizes from GPS and twin method results seem large, the GPS effect sizes are substantial given the limitations of the polygenic score method. In contrast to the twin method, which captures all types of genetic variation, GPS results are based on common DNA markers only. Furthermore, the predictive power of polygenic scores is directly related to the power of GWA studies to detect the small SNP effect sizes to begin with, which is one of the main difficulties faced in genetic research (Cesarini & Visscher, 2017). Because of a lack of statistical power attributed to sample size and other factors, such as genotyping error or measurement error of the target phenotype, effect size estimates of specific SNPs include measurement error (Dudbridge, 2003; McCarthy et al., 2008; Van Der Sluis, Verhage, Posthuma, & Dolan, 2010). Therefore, these estimates are not entirely representative of the true genetic effect, contributing to a downward bias of the GPS prediction.

Limitations

Despite the broad range of phenotypes used within the present study, there were limitations to our measures. The first limitation concerns our personality dimension reduction analysis. Although the five dimensions that emerged from this analysis were closely aligned with the literature on personality, instead of a fifth factor for Extraversion, we found a factor tapping into motivation. There are two reasons for this finding. First, the measures captured by the Academic Motivation dimension are not typically included within factor analysis of personality dimensions. These measures, (e.g., academic self-concept, self-efficacy and attitudes toward subjects) correlate with the Conscientiousness dimension ($r = 0.18–0.47$), as would be expected given its underlying facets of productivity and self-discipline; however, most of the variance is left unexplained. Second, the underlying facets of Extraversion (e.g., *gregarious*, *excitement seeking*, and *warmth*) were not well covered within our measures. For these reasons, it is not surprising that a separate factor of Extraversion did not emerge. Therefore, we excluded Extraversion from the factor analysis and used this measure by itself in an effort to maintain consistency with the wealth of existing literature describing the distinct factor structure of personality that includes Extraversion.

The second limitation with our measures was missing data. Because not everyone in our study completed all of the personality and motivation measures, there were missing data for each of our broad dimensions. To make sure that this did not affect the representativeness of the sample, we compared those with missing and nonmissing data on socioeconomic status, general cognitive ability and achievement at age 16. We found that missingness accounted for 1% to 2.6% of the variance in these outcome variables, suggesting that those with missing and nonmissing data were not substantially different on these traits.

Another limitation was that we did not have access to parental DNA. This meant that we were unable to test the effect of nontransmitted alleles that are related to years of education, neuroticism and well-being on offspring personality measures. This would make it possible to estimate the extent to which the associations between the three GPS and the personality domains are influenced by passive rGE. We were also not able to estimate potential effects of active or evocative rGE, which are difficult to investigate because of the lack of adequate measures of the environment.

A final limitation concerns a potential overfit to our data. Especially in GPS analyses where parameters for GPS construction are often chosen based on the best prediction of the outcome, prediction estimates can be inflated because of this optimization. To reduce the chance of overfit, we applied a threshold of 1 to the GPS construction, meaning that all genetic variants are retained (albeit adjusted due to linkage disequilibrium in the sample and the SNP-heritability of the trait). In a further attempt to validate our prediction estimates, we performed internal validation via repeated cross-validation as we had no access to external, independent data. We found that the more stable estimates obtained from repeated cross-validation were largely consistent with our prediction estimates, therefore indicating that our findings were comparably robust.

Conclusion

Despite the limitations to this study, it is the most comprehensive study to date investigating the link between *EduYears* GPS and personality traits. Our findings highlight the pleiotropic nature of the *EduYears* GPS and illustrate that, at a genetic level, staying in education is associated with a multitude of different traits—personality, motivation, and intelligence. Although the predictions from polygenic scores are relatively small for personality measures (between 0.6% and 2.9%), this study goes some way in starting to unpack the genetic architecture of educational achievement and associated traits, beyond what we have learnt from twin studies. As GPS prediction improves thanks to the increasing sample sizes of GWA studies and methodological advances, GPS will become more powerful for prediction of education-related measures.

References

Adams, R., & Wu, M. (Eds.). (2003). *PISA Programme for international student assessment (PISA) PISA 2000 Technical Report*. Paris, France: OECD Publishing. http://dx.doi.org/10.1787/9789264199521-en

Allegrini, A., Selzam, S., Rimfeld, K., von Stumm, S., Pingault, J.-B., & Plomin, R. (2018). Genomic prediction of cognitive traits in childhood and adolescence. *bioRxiv*. Advance online publication.

Angold, A., Costello, E. J., Messer, S. C., & Pickles, A. (1995). Development of a short questionnaire for use in epidemiological studies of depression in children and adolescents. *International Journal of Methods in Psychiatric Research, 4,* 237–249.

Appleton, J. J., Christenson, S. L., Kim, D., & Reschly, A. L. (2006). Measuring cognitive and psychological engagement: Validation of the Student Engagement Instrument. *Journal of School Psychology, 44,* 427–445. http://dx.doi.org/10.1016/j.jsp.2006.04.002

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, NY: Macmillan.

Belsky, D. W., Domingue, B. W., Wedow, R., Arseneault, L., Boardman, J. D., Caspi, A., . . . Harris, K. M. (2018). Genetic analysis of social-class mobility in five longitudinal studies. *PNAS Proceedings of the National Academy of Sciences of the United States of America, 115,* E7275–E7284. http://dx.doi.org/10.1073/pnas.1801238115

Bidjerano, T., & Dai, D. Y. (2007). The relationship between the big-five model of personality and self-regulated learning strategies. *Learning and Individual Differences, 17,* 69–81. http://dx.doi.org/10.1016/j.lindif.2007.02.001

Binet, A., & Simon, T. (1916). *The development of intelligence in children: The Binet-Simon Scale*. Baltimore, MD: Williams & Wilkins Company. http://dx.doi.org/10.1037/11069-000

Bouchard, T. J., Jr., & McGue, M. (2003). Genetic and environmental influences on human psychological differences. *Journal of Neurobiology, 54,* 4–45. http://dx.doi.org/10.1002/neu.10160

Burden, R. (1998). Assessing children's perceptions of themselves as learners and problem-solvers: The construction of the Myself-as-Learner Scale (MALS). *School Psychology International, 19,* 291–305. http://dx.doi.org/10.1177/0143034398194002

Busato, V. V., Prins, F. J., Elshout, J. J., & Hamaker, C. (1999). The relation between learning styles, the Big Five personality traits and achievement motivation in higher education. *Personality and Individual Differences, 26,* 129–140. http://dx.doi.org/10.1016/S0191-8869(98)00112-3

Busato, V. V., Prins, F. J., Elshout, J. J., & Hamaker, C. (2000). Intellectual ability, learning style, personality, achievement motivation and academic success of psychology students in higher education. *Personality and Individual Differences, 29,* 1057–1068. http://dx.doi.org/10.1016/S0191-8869(99)00253-6

Canty, A., & Ripley, B. (2012). boot: Bootstrap R (S-Plus) functions. *R package version 1.3–20*.

Caprara, G. V., Vecchione, M., Alessandri, G., Gerbino, M., & Barbaranelli, C. (2011). The contribution of personality traits and self-efficacy beliefs to academic achievement: A longitudinal study. *British Journal of Educational Psychology, 81,* 78–96. http://dx.doi.org/10.1348/2044-8279.002004

Cesarini, D., & Visscher, P. M. (2017). Genetics and educational attainment. *npj Science of Learning, 2,* 4.

Chamorro-Premuzic, T., & Furnham, A. (2003). Personality traits and academic examination performance. *European Journal of Personality, 17,* 237–250. http://dx.doi.org/10.1002/per.473

Choi, S. W., Mak, T. S. H., & O'Reilly, P. (2018). A guide to performing Polygenic Risk Score analyses. *bioRxiv*. Advance online publication.

Conard, M. A. (2006). Aptitude is not enough: How personality and behavior predict academic performance. *Journal of Research in Personality, 40,* 339–346. http://dx.doi.org/10.1016/j.jrp.2004.10.003

Deary, I. J., & Johnson, W. (2010). Intelligence and education: Causal perceptions drive analytic processes and therefore conclusions. *International Journal of Epidemiology, 39,* 1362–1369. http://dx.doi.org/10.1093/ije/dyq072

Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence, 35,* 13–21. http://dx.doi.org/10.1016/j.intell.2006.02.001

De Raad, B., & Schouwenburg, H. C. (1996). Personality in learning and education: A review. *European Journal of Personality, 10,* 303–336. http://dx.doi.org/10.1002/(SICI)1099-0984(199612)10:5<303::AID-PER262>3.0.CO;2-2

Derks, E. M., Dolan, C. V., & Boomsma, D. I. (2006). A test of the equal environment assumption (EEA) in multivariate twin studies. *Twin Research and Human Genetics, 9,* 403–411. http://dx.doi.org/10.1375/twin.9.3.403

Duckworth, A. L., & Quinn, P. D. (2009). Development and validation of the short grit scale (grit-s). *Journal of Personality Assessment, 91,* 166–174. http://dx.doi.org/10.1080/00223890802634290

Dudbridge, F. (2003). Pedigree disequilibrium tests for multilocus haplotypes. *Genetic Epidemiology, 25,* 115–121. http://dx.doi.org/10.1002/gepi.10252

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research, 74,* 59–109. http://dx.doi.org/10.3102/00346543074001059

Frey, M. C., & Detterman, D. K. (2004). Scholastic Assessment or g? The relationship between the Scholastic Assessment Test and general cognitive ability. *Psychological Science, 15,* 373–378. http://dx.doi.org/10.1111/j.0956-7976.2004.00687.x

Geramian, S. M., Mashayekhi, S., & Ninggal, M. T. B. H. (2012). The relationship between personality traits of international students and academic achievement. *Procedia: Social and Behavioral Sciences, 46,* 4374–4379. http://dx.doi.org/10.1016/j.sbspro.2012.06.257

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology & Psychiatry, & Allied Disciplines, 38,* 581–586. http://dx.doi.org/10.1111/j.1469-7610.1997.tb01545.x

Gratten, J., Wray, N. R., Keller, M. C., & Visscher, P. M. (2014). Large-scale genomics unveils the genetic architecture of psychiatric disorders. *Nature Neuroscience, 17,* 782–790. http://dx.doi.org/10.1038/nn.3708

Haworth, C. M., Davis, O. S., & Plomin, R. (2013). Twins Early Development Study (TEDS): A genetically sensitive investigation of cognitive and behavioral development from childhood to young adulthood. *Twin Research and Human Genetics, 16,* 117–125. http://dx.doi.org/10.1017/thg.2012.91

Heaven, P. C., & Ciarrochi, J. (2008). Parental styles, conscientiousness, and academic performance in high school: A three-wave longitudinal study. *Personality and Social Psychology Bulletin, 34,* 451–461. http://dx.doi.org/10.1177/0146167207311909

Hirschberg, N., & Itkin, S. (1978). Graduate student success in psychology. *American Psychologist, 33,* 1083–1093. http://dx.doi.org/10.1037/0003-066X.33.12.1083

Holland, D. C., Dollinger, S. J., Holland, C. J., & MacDonald, D. A. (1995). The relationship between psychometric intelligence and the five-factor model of personality in a rehabilitation sample. *Journal of Clinical Psychology, 51,* 79–88. http://dx.doi.org/10.1002/1097-4679(199501)51:1<79::AID-JCLP2270510113>3.0.CO;2-P

Huebner, E. S. (1994). Preliminary development and validation of a multidimensional life satisfaction scale for children. *Psychological Assessment, 6,* 149–158. http://dx.doi.org/10.1037/1040-3590.6.2.149

Jang, K. L., Livesley, W. J., & Vemon, P. A. (1996). Heritability of the big five personality dimensions and their facets: A twin study. *Journal of Personality, 64,* 577–592. http://dx.doi.org/10.1111/j.1467-6494.1996.tb00522.x

Jencks, C. (1979). *Who gets ahead? The determinants of economic success in America*. New York, NY: Basic Books.

Kaplan, E., Fein, D., Kramer, J., Delis, D., & Morris, R. (1999). *The WISC-III as a process instrument*. San Antonio, TX: The Psychological Corporation.

Kashdan, T. B., Rose, P., & Fincham, F. D. (2004). Curiosity and exploration: Facilitating positive subjective experiences and personal growth opportunities. *Journal of Personality Assessment, 82,* 291–305. http://dx.doi.org/10.1207/s15327752jpa8203_05

Kelly, W. E., & Johnson, J. L. (2005). Time use efficiency and the five-factor model of personality. *Education, 125,* 511–515.

Kendler, K. S., Neale, M. C., Kessler, R. C., Heath, A. C., & Eaves, L. J. (1993). A test of the equal-environment assumption in twin studies of psychiatric illness. *Behavior Genetics, 23,* 21–27. http://dx.doi.org/10.1007/BF01067551

Kim, J.-H. (2009). Estimating classification error rate: Repeated cross-validation, repeated hold-out and bootstrap. *Computational Statistics & Data Analysis, 53,* 3735–3745. http://dx.doi.org/10.1016/j.csda.2009.04.009

Knopik, V. S., Neiderhiser, J. M., DeFries, J. C., & Plomin, R. (2017). *Behavioral genetics* (7th ed.). London, UK: Worth Publishers.

Komarraju, M., & Karau, S. J. (2005). The relationship between the big five personality traits and academic motivation. *Personality and Individual Differences, 39,* 557–567. http://dx.doi.org/10.1016/j.paid.2005.02.013

Kong, A., Thorleifsson, G., Frigge, M. L., Vilhjalmsson, B. J., Young, A. I., Thorgeirsson, T. E., . . . Stefansson, K. (2018). The nature of nurture: Effects of parental genotypes. *Science, 359,* 424–428. http://dx.doi.org/10.1126/science.aan6877

Krapohl, E., Rimfeld, K., Shakeshaft, N. G., Trzaskowski, M., McMillan, A., Pingault, J.-B., . . . Plomin, R. (2014). The high heritability of educational achievement reflects many genetically influenced traits, not just intelligence. *PNAS Proceedings of the National Academy of Sciences of the United States of America, 111,* 15273–15278. http://dx.doi.org/10.1073/pnas.1408777111

Kuhn, M. (2015). caret: Classification and regression training. *Astrophysics Source Code Library*.

Laidra, K., Pullmann, H., & Allik, J. (2007). Personality and intelligence as predictors of academic achievement: A cross-sectional study from elementary to secondary school. *Personality and Individual Differences, 42,* 441–451. http://dx.doi.org/10.1016/j.paid.2006.08.001

Lakes, K. D., Swanson, J. M., & Riggs, M. (2012). The reliability and validity of the English and Spanish Strengths and Weaknesses of ADHD and Normal behavior rating scales in a preschool sample: Continuum measures of hyperactivity and inattention. *Journal of Attention Disorders, 16,* 510–516. http://dx.doi.org/10.1177/1087054711413550

Lee, J. J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., . . . the 23andMe Research Team, & the COGENT (Cognitive Genomics Consortium), & the Social Science Genetic Association Consortium. (2018). Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature Genetics, 50,* 1112–1121. http://dx.doi.org/10.1038/s41588-018-0147-3

Luciano, M., Hagenaars, S. P., Davies, G., Hill, W. D., Clarke, T.-K., Shirali, M., . . . Deary, I. J. (2018). Association analysis in over 329,000 individuals identifies 116 independent variants influencing neuroticism. *Nature Genetics, 50,* 6–11. http://dx.doi.org/10.1038/s41588-017-0013-8

Lyubomirsky, S., & Lepper, H. S. (1999). A measure of subjective happiness: Preliminary reliability and construct validation. *Social Indicators Research, 46,* 137–155. http://dx.doi.org/10.1023/A:1006824100041

Mason, O., Linney, Y., & Claridge, G. (2005). Short scales for measuring schizotypy. *Schizophrenia Research, 78,* 293–296. http://dx.doi.org/10.1016/j.schres.2005.06.020

McCarthy, D. (1972). *McCarthy scales of children's abilities*. New York, NY: Psychological Corporation.

McCarthy, M. I., Abecasis, G. R., Cardon, L. R., Goldstein, D. B., Little, J., Ioannidis, J. P., & Hirschhorn, J. N. (2008). Genome-wide association studies for complex traits: Consensus, uncertainty and challenges. *Nature Reviews Genetics, 9,* 356–369. http://dx.doi.org/10.1038/nrg2344

McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A. R., Teumer, A., . . . the Haplotype Reference Consortium. (2016). A reference panel of 64,976 haplotypes for genotype imputation. *Nature Genetics, 48,* 1279–1283. http://dx.doi.org/10.1038/ng.3643

McCrae, R. R., & Costa, P. T. (1997). Conceptions and correlates of openness to experience. In R. Hogan, J. Johnson, & S. Briggs (Eds.), *Handbook of personality psychology* (pp. 825–847). Cambridge, MA: Elsevier. http://dx.doi.org/10.1016/B978-012134645-4/50032-9

McCrae, R. R., & Costa, P. T., Jr. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology, 52,* 81–90. http://dx.doi.org/10.1037/0022-3514.52.1.81

McCullough, M. E., Emmons, R. A., & Tsang, J.-A. (2002). The grateful disposition: A conceptual and empirical topography. *Journal of Personality and Social Psychology, 82,* 112–127. http://dx.doi.org/10.1037/0022-3514.82.1.112

Mieloo, C., Raat, H., van Oort, F., Bevaart, F., Vogel, I., Donker, M., & Jansen, W. (2012). Validity and reliability of the strengths and difficulties questionnaire in 5–6 year olds: Differences by gender or by parental education? *PLoS ONE, 7,* e36805. http://dx.doi.org/10.1371/journal.pone.0036805

Mõttus, R., Realo, A., Vainik, U., Allik, J., & Esko, T. (2017). Educational attainment and personality are genetically intertwined. *Psychological Science, 28,* 1631–1639. http://dx.doi.org/10.1177/0956797617719083

Mullins-Sweatt, S. N., Jamerson, J. E., Samuel, D. B., Olson, D. R., & Widiger, T. A. (2006). Psychometric properties of an abbreviated instrument of the five-factor model. *Assessment, 13,* 119–137. http://dx.doi.org/10.1177/1073191106286748

Multon, K. D., Brown, S. D., & Lent, R. W. (1991). Relation of self-efficacy beliefs to academic outcomes: A meta-analytic investigation. *Journal of Counseling Psychology, 38,* 30–38. http://dx.doi.org/10.1037/0022-0167.38.1.30

Nyholt, D. R. (2004). A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. *American Journal of Human Genetics, 74,* 765–769. http://dx.doi.org/10.1086/383251

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and velicer's MAP test. *Behavior Research Methods, Instruments, & Computers, 32,* 396–402. http://dx.doi.org/10.3758/BF03200807

O'Connor, M. C., & Paunonen, S. V. (2007). Big Five personality predictors of post-secondary academic performance. *Personality and Individual Differences, 43,* 971–990. http://dx.doi.org/10.1016/j.paid.2007.03.017

OECD. (2007). *Education at a glance 2007*. Paris, France: Organisation for Economic Cooperation and Development.

OECD. (2017). *Education at a glance 2017*. Paris, France: Organisation for Economic Cooperation and Development.

Okbay, A., Baselmans, B. M., De Neve, J.-E., Turley, P., Nivard, M. G., Fontana, M. A., . . . the LifeLines Cohort Study. (2016). Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nature Genetics, 48,* 624–633. http://dx.doi.org/10.1038/ng.3552

Okbay, A., Beauchamp, J. P., Fontana, M. A., Lee, J. J., Pers, T. H., Rietveld, C. A., . . . the LifeLines Cohort Study. (2016). Genome-wide association study identifies 74 loci associated with educational attainment. *Nature, 533,* 539–542. http://dx.doi.org/10.1038/nature17671

Parker, P. D., Marsh, H. W., Ciarrochi, J., Marshall, S., & Abduljabbar, A. S. (2014). Juxtaposing math self-efficacy and self-concept as predic-

tors of long-term achievement outcomes. *Educational Psychology, 34,* 29–48. http://dx.doi.org/10.1080/01443410.2013.797339

Petrides, K. V., Chamorro-Premuzic, T., Frederickson, N., & Furnham, A. (2005). Explaining individual differences in scholastic behaviour and achievement. *British Journal of Educational Psychology, 75,* 239–255. http://dx.doi.org/10.1348/000709904X24735

Plomin, R., & Deary, I. J. (2015). Genetics and intelligence differences: Five special findings. *Molecular Psychiatry, 20,* 98–108. http://dx.doi.org/10.1038/mp.2014.105

Plomin, R., DeFries, J. C., & Loehlin, J. C. (1977). Genotype-environment interaction and correlation in the analysis of human behavior. *Psychological Bulletin, 84,* 309–322. http://dx.doi.org/10.1037/0033-2909.84.2.309

Plomin, R., & von Stumm, S. (2018). The new genetics of intelligence. *Nature Reviews Genetics, 19,* 148–159. http://dx.doi.org/10.1038/nrg.2017.104

Polderman, T. J., Benyamin, B., de Leeuw, C. A., Sullivan, P. F., van Bochoven, A., Visscher, P. M., & Posthuma, D. (2015). Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nature Genetics, 47,* 702–709. http://dx.doi.org/10.1038/ng.3285

Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological Bulletin, 135,* 322–338. http://dx.doi.org/10.1037/a0014996

Purcell, S. M., Wray, N. R., Stone, J. L., Visscher, P. M., O'Donovan, M. C., Sullivan, P. F., . . . the International Schizophrenia Consortium. (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature, 460,* 748–752. http://dx.doi.org/10.1038/nature08185

Raiche, G., & Magis, D. (2010). nFactors: An R package for parallel analysis and non graphical solutions to the Cattell scree test. *R package version 2,* 2.

Raven, J., Raven, J. C., & Court, J. H. (1998). *Manual for Raven's progressive matrices and vocabulary scales*. Oxford, UK: Oxford Psychologists.

Raven, J. C., Raven, J. E., & Court, J. H. (1989). *Mill Hill vocabulary scale*. San Antonio, TX: Psychological Corporation.

R Core Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: Foundation for Statistical Computing.

Revelle, W. R. (2017). *psych: Procedures for Personality and Psychological Research*. Chicago, IL: Northwestern University Scholars.

Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin, 138,* 353–387. http://dx.doi.org/10.1037/a0026838

Rietveld, C. A., Medland, S. E., Derringer, J., Yang, J., Esko, T., Martin, N. W., . . . the LifeLines Cohort Study. (2013). GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science, 340,* 1467–1471. http://dx.doi.org/10.1126/science.1235488

Ritchie, S. J., & Tucker-Drob, E. M. (2018). How much does education improve intelligence? A meta-analysis. *Psychological Science, 29,* 1358–1369. http://dx.doi.org/10.1177/0956797618774253

Rosseel, Y., Oberski, D., Byrnes, J., Vanbrabant, L., Savalei, V., Merkle, E., . . . Chow, M. (2018). The lavaan package for R.

Scheier, M. F., Carver, C. S., & Bridges, M. W. (1994). Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): A reevaluation of the Life Orientation Test. *Journal of Personality and Social Psychology, 67,* 1063–1078. http://dx.doi.org/10.1037/0022-3514.67.6.1063

Schunk, D. H. (1989). Self-efficacy and achievement behaviors. *Educational Psychology Review, 1,* 173–208. http://dx.doi.org/10.1007/BF01320134

Selzam, S., Dale, P. S., Wagner, R. K., DeFries, J. C., Cederlöf, M., O'Reilly, P. F., . . . Plomin, R. (2017). Genome-wide polygenic scores predict reading performance throughout the school years. *Scientific Studies of Reading, 21,* 334–349. http://dx.doi.org/10.1080/10888438.2017.1299152

Selzam, S., Krapohl, E., von Stumm, S., O'Reilly, P. F., Rimfeld, K., Kovas, Y., . . . Plomin, R. (2017). Predicting educational achievement from DNA. *Molecular Psychiatry, 22,* 267–272. http://dx.doi.org/10.1038/mp.2016.107

Selzam, S., McAdams, T. A., Coleman, J. R. I., Carnell, S., O'Reilly, P. F., Plomin, R., & Llewellyn, C. H. (2018). Evidence for gene-environment correlation in child feeding: Links between common genetic variation for BMI in children and parental feeding practices. *PLOS Genetics, 14,* e1007757. http://dx.doi.org/10.1371/journal.pgen.1007757

Šidák, Z. (1971). On probabilities of rectangles in multivariate Student distributions: Their dependence on correlations. *Annals of Mathematical Statistics, 42,* 169–175. http://dx.doi.org/10.1214/aoms/1177693504

Silverman, W. K., Fleisig, W., Rabian, B., & Peterson, R. A. (1991). Childhood anxiety sensitivity index. *Journal of Clinical Child Psychology, 20,* 162–168. http://dx.doi.org/10.1207/s15374424jccp2002_7

Snyder, C. R., Hoza, B., Pelham, W. E., Rapoff, M., Ware, L., Danovsky, M., . . . Stahl, K. J. (1997). The development and validation of the Children's Hope Scale. *Journal of Pediatric Psychology, 22,* 399–421. http://dx.doi.org/10.1093/jpepsy/22.3.399

Socrates, A., Bond, T., Karhunen, V., Auvinen, J., Rietveld, C., Veijola, J . . . . O'Reilly, P. (2017). Polygenic risk scores applied to a single cohort reveal pleiotropy among hundreds of human phenotypes. *bioRxiv*. Advance online publication.

Solovieff, N., Cotsapas, C., Lee, P. H., Purcell, S. M., & Smoller, J. W. (2013). Pleiotropy in complex traits: Challenges and strategies. *Nature Reviews Genetics, 14,* 483–495. http://dx.doi.org/10.1038/nrg3461

Spinath, B., Spinath, F. M., Harlaar, N., & Plomin, R. (2006). Predicting school achievement from general cognitive ability, self-perceived ability, and intrinsic value. *Intelligence, 34,* 363–374. http://dx.doi.org/10.1016/j.intell.2005.11.004

Swanson, J. M., Schuck, S., Porter, M. M., Carlson, C., Hartman, C. A., Sergeant, J. A., . . . Wigal, T. (2012). Categorical and Dimensional Definitions and Evaluations of Symptoms of ADHD: History of the SNAP and the SWAN Rating Scales. *The International Journal of Educational and Psychological Assessment, 10,* 51–70.

Trautwein, U., & Lüdtke, O. (2007). Students' self-reported effort and time on homework in six school subjects: Between-students differences and within-student variation. *Journal of Educational Psychology, 99,* 432–444. http://dx.doi.org/10.1037/0022-0663.99.2.432

Trautwein, U., Lüdtke, O., Roberts, B. W., Schnyder, I., & Niggli, A. (2009). Different forces, same consequence: Conscientiousness and competence beliefs are independent predictors of academic effort and achievement. *Journal of Personality and Social Psychology, 97,* 1115–1128. http://dx.doi.org/10.1037/a0017048

Trzaskowski, M., Eley, T. C., Davis, O. S., Doherty, S. J., Hanscombe, K. B., Meaburn, E. L., . . . Plomin, R. (2013). First genome-wide association study on anxiety-related behaviours in childhood. *PLoS ONE, 8,* e58676. http://dx.doi.org/10.1371/journal.pone.0058676

Tucker-Drob, E. M., Briley, D. A., Engelhardt, L. E., Mann, F. D., & Harden, K. P. (2016). Genetically-mediated associations between measures of childhood character and academic achievement. *Journal of Personality and Social Psychology, 111,* 790–815. http://dx.doi.org/10.1037/pspp0000098

van der Sluis, S., Verhage, M., Posthuma, D., & Dolan, C. V. (2010). Phenotypic complexity, measurement bias, and poor phenotypic resolution contribute to the missing heritability problem in genetic association studies. *PLoS ONE, 5,* e13929. http://dx.doi.org/10.1371/journal.pone.0013929

Vilhjálmsson, B. J., Yang, J., Finucane, H. K., Gusev, A., Lindström, S., Ripke, S., . . . Price, A. (2015). Modeling linkage disequilibrium in-

creases accuracy of polygenic risk scores. *American Journal of Human Genetics, 97,* 576–592.

Wagerman, S. A., & Funder, D. C. (2007). Acquaintance reports of personality and academic achievement: A case for conscientiousness. *Journal of Research in Personality, 41,* 221–229. http://dx.doi.org/10.1016/j.jrp.2006.03.001

Webb, E. (1915). *Character and intelligence: An attempt at an exact study of character*. London, UK: University Press.

Wechsler, D., Golombok, S., & Rust, J. (1992). *WISC-III U. K. Wechsler intelligence scale for children: U. K. manual*. Sidcup, UK: The Psychological Corporation.

Williams, E. J. (1959). The comparison of regression variables. *Journal of the Royal Statistical Society: Series A (General), 21,* 396–399.

Wray, N. R., Yang, J., Hayes, B. J., Price, A. L., Goddard, M. E., & Visscher, P. M. (2013). Pitfalls of predicting complex traits from SNPs. *Nature Reviews Genetics, 14,* 507–515. http://dx.doi.org/10.1038/nrg3457

Zeidner, M., & Matthews, G. (2000). Intelligence and personality. In R. J. Sternberg (Ed.), *Handbook of intelligence* (pp. 581–610). Cambridge, UK: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511807947.027

Zeileis, A., & Hothorn, T. (2002). Diagnostic checking in regression relationships. *R News, 2,* 7–10.

Received November 13, 2017
Revision received January 4, 2019
Accepted January 5, 2019 ■

Chapter 4 – A polygenic *p* factor for major psychiatric disorders

This chapter is presented as a published paper. It is an exact copy of this publication:

Selzam, S., Coleman, J. R., Caspi, A., Moffitt, T. E., & Plomin, R. (2018). A polygenic p factor for major psychiatric disorders. *Translational psychiatry*, 8(1), 205. doi: 10.1038/s41398-018-0217-4

Supplemental materials are included under Appendix 3.

Selzam et al. *Translational Psychiatry* (2018)8:205
DOI 10.1038/s41398-018-0217-4

Translational Psychiatry

ARTICLE Open Access

A polygenic *p* factor for major psychiatric disorders

Saskia Selzam[1], Jonathan R. I. Coleman[1,2], Avshalom Caspi[1,3,4,5], Terrie E. Moffitt[1,3,4,5] and Robert Plomin[1]

Abstract

It has recently been proposed that a single dimension, called the *p* factor, can capture a person's liability to mental disorder. Relevant to the *p* hypothesis, recent genetic research has found surprisingly high genetic correlations between pairs of psychiatric disorders. Here, for the first time, we compare genetic correlations from different methods and examine their support for a genetic *p* factor. We tested the hypothesis of a genetic *p* factor by applying principal component analysis to matrices of genetic correlations between major psychiatric disorders estimated by three methods—family study, genome-wide complex trait analysis, and linkage-disequilibrium score regression—and on a matrix of polygenic score correlations constructed for each individual in a UK-representative sample of 7 026 unrelated individuals. All disorders loaded positively on a first unrotated principal component, which accounted for 57, 43, 35, and 22% of the variance respectively for the four methods. Our results showed that all four methods provided strong support for a genetic *p* factor that represents the pinnacle of the hierarchical genetic architecture of psychopathology.

Introduction

High comorbidity rates among psychiatric disorders[1] have led to research investigating higher-order dimensions for psychopathology, including Internalizing (e.g., Anxiety and Depression), Externalizing (e.g., Hyperactivity and Conduct Disorder), and Psychotic Experiences (e.g., Schizophrenia and Bipolar Disorder)[2]. However, these higher-order dimensions also correlate with each other[3], which suggests the possible existence of a general factor of psychopathology[4]. This general factor has been called the *p* factor[5] as it captures the shared variance across psychiatric symptoms, and predicts a multitude of poor outcomes and general life impairment[6,7].

Family studies support the hypothesis of a genetic *p* factor in that genetic influences on psychopathology appear to be general across disorders rather than specific to each disorder. For example, psychiatric disorders do not breed true—parental psychopathology predicts offspring psychiatric disorders but with little specificity[8]. Family research has found substantial genetic correlations between pairs of disorders, such as Major Depression and Generalized Anxiety Disorder[9] and Schizophrenia and Bipolar Disorder[10]. Genetic overlap between internalizing and externalizing higher-order constructs has also been noted[11], consistent with the hypothesis of a general *p* factor. The culmination of this research is a recent study of more than 3 million full and half-siblings using Swedish national register data that found evidence for a general genetic factor that pervades eight major psychiatric disorders as well as convictions for violent crimes[12]. Although genetic correlations were not presented, the average loading was 0.45 on a general genetic factor.

Genomic research also supports the hypothesis of a genetic *p* factor. The first hint came from genome-wide association (GWA) findings that single- nucleotide polymorphisms (SNPs) found to be associated with Schizophrenia were also associated with bipolar disorder[13]. In 2013, genetic correlations were first estimated from linear mixed model analyses (genome-wide complex trait

Correspondence: Saskia Selzam (saskia.selzam@kcl.ac.uk)
[1]MRC Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK
[2]NIHR Biomedical Research Centre for Mental Health, South London and Maudsley NHS Trust, London, UK
Full list of author information is available at the end of the article.

analysis, GCTA) of individual genotype data for five psychiatric disorders in the Psychiatric Genomics Consortium (PGC)[14]. Schizophrenia, Bipolar Disorder and Major Depressive Disorder yielded the highest genetic intercorrelations (average = 0.53); the average genetic correlation among the five disorders, including Autistic Spectrum Disorder and Attention-Deficit/Hyperactivity Disorder, was 0.22.

Linkage-Disequilibrium Score Regression (LDSC)[15] has made it possible to estimate genetic correlations from GWA summary statistics rather than requiring genotype data for individuals. This method is based on correlations in effect sizes across disorders taking into account linkage disequilibrium and the SNP heritabilities of the disorders. LDSC genetic correlations derived from summary GWA statistics for the same five PGC disorders are remarkably similar to the GCTA genetic correlations described above that used individual genotype data[16]. A recent LDSC analysis of eight psychiatric disorders again showed considerable correlations between Schizophrenia, Bipolar Disorder and Major Depressive Disorder (average = 0.41), and yielded an average genetic correlation of 0.21[17], highlighting the relevance of testing the hypothesis of a genetic *p* factor.

Another approach that has not yet been systematically applied to test for a genetic *p* is to correlate genome-wide polygenic scores (GPS), although some GPS correlations between pairs of psychiatric disorders have been reported[18]. A GPS for a disorder is created for an individual by summing the count of alleles shown in GWA studies to be associated with the disorder, after weighting the alleles by the strength of their association[19]. The previously described PGC dataset was used to create polygenic scores for each of the five disorders[13], and polygenic scores for Schizophrenia, Bipolar Disorder and Major Depressive Disorder predicted liability variance in the other disorders, again suggesting genetic overlap. However, as new GWA studies have been published since for Schizophrenia, Attention-Deficit/Hyperactivity Disorder and Autism Spectrum Disorder with considerably increased sample sizes, replication is needed. GPS correlations between disorders are related to genetic correlations, but differ from the genetic correlations estimated from other methods because they index both the relationship between individual-specific genetic effects for traits in the population and genetic effects derived from an independent analysis. Nonetheless, GPS correlations provide another opportunity to test the hypothesis of a genetic *p* factor.

Based on the overwhelming evidence that favors a general *p* factor, we test whether a general *p* factor also emerges from genomic data. In the present study, we bring together genetic correlations for major psychiatric disorders derived from four genetic methods (family, GCTA, LDSC and GPS). We applied principal component analysis to correlation matrices derived from these four methods and estimate the amount of genetic variance explained by a genetic *p* factor. For the GPS approach, we constructed GPS for eight psychiatric disorders for each individual in a sample of 7 026 unrelated individuals from the Twins Early Development Study (TEDS)[20].

Our hypothesis was that a general genetic factor would emerge from factor analyses of correlations derived from each of the four genetic methods. We also investigated the extent to which all disorders load on this general factor and the magnitude of their loadings.

Methods

Sample

This study included 7 026 unrelated (i.e., one member per twin pair), genotyped individuals from TEDS, a longitudinal birth cohort that recruited over 1 5000 twin pairs between 1994–1996 who were born in England or Wales. Despite some attrition, the remaining cohort, as well as the genotyped subsample have been shown to represent the UK population[20,21]. Written informed consent was obtained from parents. Project approval was granted by King's College London's ethics committee for the Institute of Psychiatry, Psychology and Neuroscience (05.Q0706/228).

GPS calculation and GPS correlations

To obtain individual-specific genetic measures for psychiatric traits, we created eight GPS in our independent sample of 7026 individuals based on publicly available GWA summary statistics from the PGC: Schizophrenia, Bipolar Disorder, Major Depressive Disorder, Autism Spectrum Disorder, Attention-Deficit/Hyperactivity Disorder, Obsessive-Compulsive Disorder, Anorexia Nervosa, Post-Traumatic Stress Disorder (Supplementary Table S1). Following quality control and imputation (see Supplementary Methods S1 for details), genotypic data included 515 100 genotyped or imputed SNPs (info = 1). To calculate polygenic scores, we used a Bayesian approach, *LDpred*[22], which modifies the summary statistic coefficients based on information on linkage disequilibrium (LD) and a prior on the effect size of each SNP. The final GPS is obtained as the sum of the trait-increasing alleles (each variant coded as 0, 1 or 2), weighted by the posterior effect size estimates. For our analyses, we used a prior that assumes a fraction of causal markers of 1 (for more information, see Supplementary Methods S2). All polygenic scores were adjusted for the first ten principal components of the genotype data, and chip, batch and plate effects using the regression method. The resulting standardized residuals were used for subsequent analyses.

In the TEDS sample, we created polygenic scores for the eight psychopathology traits. These scores followed a normal distribution and were used to generate a

correlation matrix for these eight polygenic scores for use in subsequent analyses.

Genetic correlations based on LDSC

LDSC is a method used to estimate SNP-heritability (SNP$-h^2$) based on GWA summary statistics only, and relies on the principle that the presence of LD in the study sample is correlated with the upward bias of GWA test statistics[15]. Cross-trait LDSC[16] is an extension of this method and makes it possible to estimate the genetic relationship between two traits. For each SNP, this method establishes the covariance of the test statistics for trait x and trait y, and regresses this value on the LD score of that SNP (i.e., the sum of the squared correlations of the SNP with its surrounding SNPs), whereby the slope represents the genetic covariance. The genetic correlation is obtained by standardizing the covariance by the SNP-h^2 for both traits ($r_g = cov_{xy}/\sqrt{h_{x^2}h_{y^2}}$). We applied cross-trait LDSC analysis on the same eight PGC summary statistics used for polygenic score creation to generate a genetic correlation matrix for further analysis. (For univariate SNP-h^2 results using LDSC, see Supplementary Table S2.)

Genetic correlations based on GCTA

In addition to GPS and LDSC analysis, we also obtained genetic correlation matrices through cross-sample bivariate GCTA based on genome-wide relatedness maximum likelihood[23]. Unlike LDSC, which uses GWA summary statistics, bivariate GCTA requires individual-level genotype data of unrelated individuals to estimate genetic correlations, implementing linear mixed model analysis. Cross-sample GCTA is an extension to bivariate GCTA[24] and makes it possible to calculate genetic correlation estimates without requiring overlapping phenotypic information between samples. Rather, it compares genetic similarity between individuals that have the same disease status (case, control) for different disorders. For example, if cases of one disorder are genetically more similar to cases of a different disorder than to the respective controls, a positive genetic correlation can be inferred. For this study, we used published cross-sample GCTA genetic correlations[14], which included five psychiatric disorders: Schizophrenia, Bipolar Disorder, Major Depressive Disorder, Autism Spectrum Disorder, and Attention-Deficit/Hyperactivity Disorder. (For univariate SNP-h^2 estimates, see Supplementary Table S3.)

Genetic correlations based on family data

Finally, we used genetic correlations based on quantitative genetic analysis comparing 3 475 122 Swedish full-siblings and half-siblings, who are genetically similar 50 and 25%, respectively, for additive genetic effects. This family study represents a very different methodology as compared to the other methods. Rather than using direct estimates based on DNA differences, it uses indirect estimates based on the relative resemblance of full siblings and half siblings. Because this family study, the only one of its kind, is so different from the other methods, it is especially valuable to compare its genetic correlations to those from the other three methods. The genetic correlations were not included in the original publication[12] but were kindly prepared and shared by the lead author, Erik Pettersson of the Karolinska Institute. The analysis included seven psychopathology traits (Schizophrenia, Bipolar Disorder, Attention-Deficit/Hyperactivity Disorder, Major Depressive Disorder, Anxiety, Alcohol use Disorder and Drug Abuse), as well as convictions for Violent Crimes. Schizoaffective disorder was redundant with Schizophrenia (genetic correlation = 0.99) and thus omitted here (Supplementary Figure S1).

Statistical analyses

Principal component analysis

To test the hypothesis that a general genetic p factor emerges from the genetic relationships among psychopathology traits, we performed eigenvalue decomposition through principal component analysis (PCA), which aims to maximize variation of the first principal component[25]. We applied PCA to genetic correlation matrices derived from family analysis (8 × 8 matrix), GCTA (5 × 5 matrix), LDSC (8 × 8 matrix), and GPS (8 × 8 matrix) to estimate the loadings of each psychiatric trait on this component and the variance explained by the first principal component.

We also tested the statistical significance of the factor loadings, which represent correlations between the original standardized variables and the factors. By calculating the t-statistic of the correlation coefficients, we were able to derive empirical p-values based on the t-statistic distribution with $n-2$ degrees of freedom[26]. Significance testing was applied only to family and GPS loadings because we were unable to obtain degrees of freedom for GCTA and LDSC data, which is required for the calculation of t. All tests were two-tailed and a significance level of $\alpha = 0.05$ was accepted as statistically significant. In addition to testing statistical significance, we calculated the proportion of factor loadings with a magnitude of $\geq|0.30|$. This value is a commonly used threshold in the factor analysis literature, as it indicates that the factor explains ~10% of the variance in the measure[27], therefore substantially contributing to the factor.

The decision of how many components to retain for rotation was based on three criteria: (i) the Kaiser criterion[28] of eigenvalue $\lambda > 1$; (ii) parallel analysis[29], and (iii) scree plot inspection[30] (for a more detailed description, see Supplementary Methods S3). To improve interpretability of the extracted components, we performed oblique rotation using the *Oblimin* method. We chose this approach, which permits factors to be correlated, because previous work

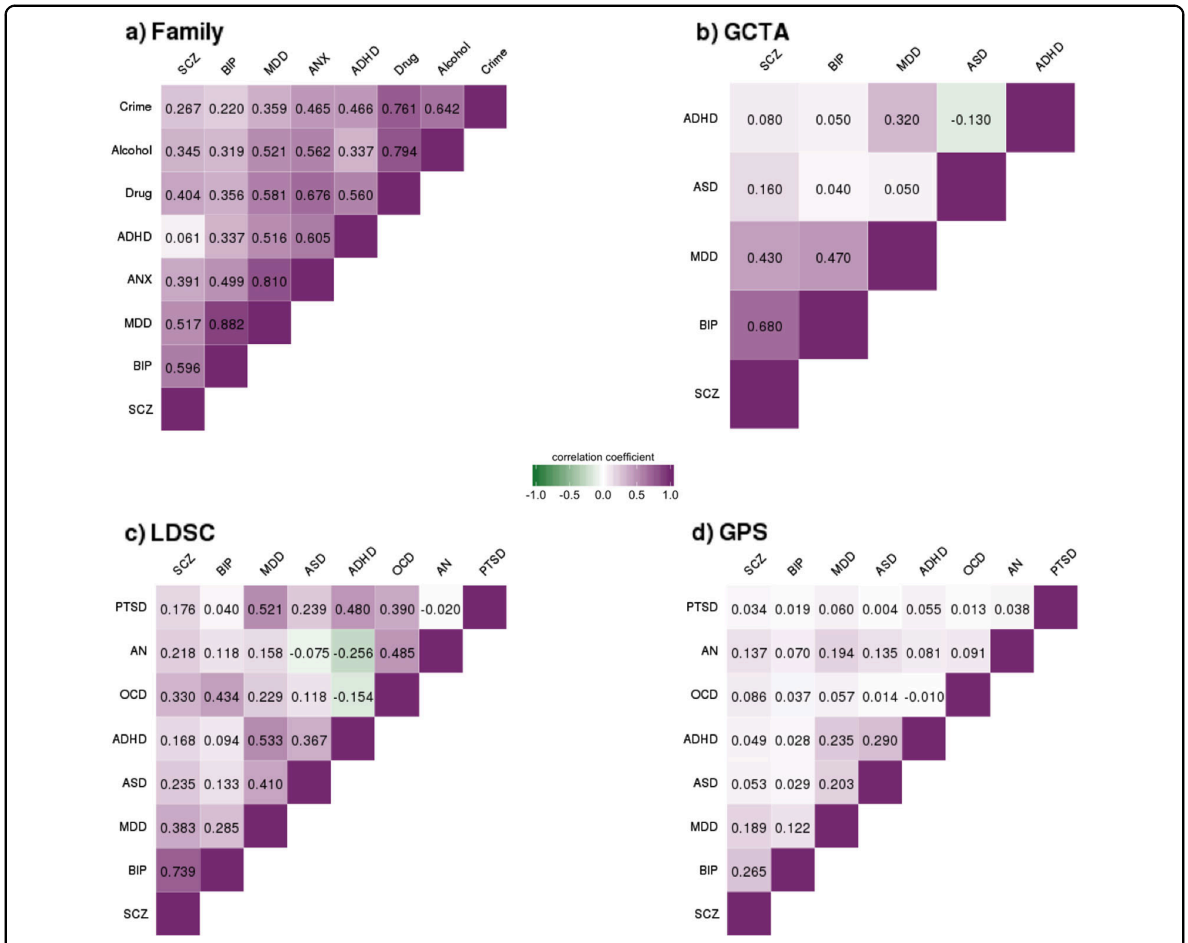


Fig. 1 Genetic correlations from family analysis (**a**), Genome-wide Complex Trait Analysis (**b**), Linkage-Disequilibrium Score Regression (**c**) and Genome-wide Polygenic Score (**d**) analysis. Values represent genetic correlations for (**a**), (**b**) and (**c**) and Pearson's correlation coefficients for (**d**). *SCZ* Schizophrenia, *BIP* Bipolar Disorder, *MDD* Major Depressive Disorder, *ASD* Autism Spectrum Disorder, *ADHD* Attention-Deficit/Hyperactivity Disorder, *ANX* Anxiety, *OCD* Obsessive-Compulsive Disorder, *AN* Anorexia Nervosa, *PTSD* Post-Traumatic Stress Disorder; Drug = Drug Abuse; Alcohol = Alcohol Abuse; Crime = Convictions of Violent Crimes

using phenotypic data showed considerable associations between latent psychopathology dimensions[3,5].

Analyses were performed in the open-source software R[31], using the *hornpa*[32] package to perform parallel analysis, the *psych*[33] package to conduct PCA (using the 'principal' function), and the *GPArotation*[34] package to apply oblique rotation. Analysis scripts are available from the first author upon request.

Results

Genetic correlations

Figure 1 presents the genetic correlations from family analysis, GCTA and LDSC, and the correlations from GPS analysis. The average genetic correlations were 0.49 for family analysis, 0.22 for GCTA and 0.24 for LDSC, indicating general genetic overlap among psychiatric disorders. The average GPS correlation was lower (0.09), as expected. However, correlations for all four genetic approaches clustered in a strikingly similar way. Most notably, the average genetic correlations between Schizophrenia, Bipolar and Depression were consistently the largest in magnitude −0.67 for family analysis, 0.53 for GCTA, 0.47 for LDSC, and 0.19 for GPS. High genetic correlations were not driven by larger heritability estimates for these traits in comparison to the other disorders (see Supplementary Tables S2 and S3 for SNP-h^2 estimates).

Principal component analysis

PCA provided converging evidence for a general psychopathology factor. Figure 2 shows that all four correlation matrices yielded first unrotated principal components with larger eigenvalues than the subsequent

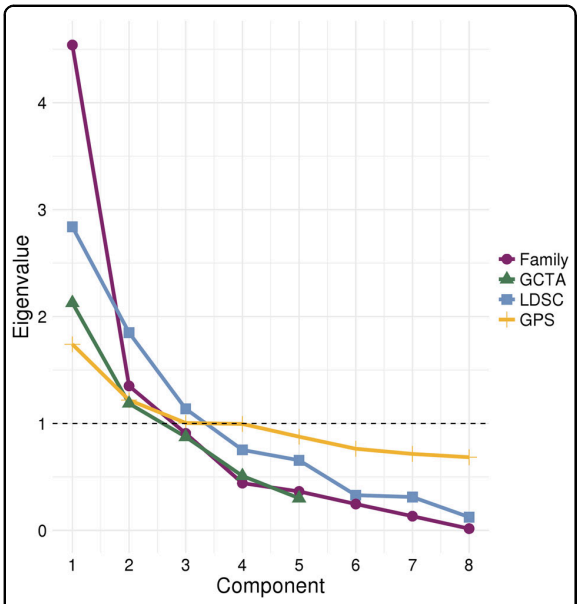


Fig. 2 Scree plot showing eigenvalues for each principal component after performing PCA on correlation matrices for four genetically sensitive methods: family analysis, Genome-wide Complex Trait Analysis (GCTA), Linkage-Disequilibrium Score Regression (LDSC) and Genome-wide Polygenic Scoring (GPS). The dashed line represents the cut-off for principal component retention based on the Kaiser's $\lambda > 1$ criterion[28]

components. The first principal component accounted for 57, 43, 35 and 22% in family, GCTA, LDSC and GPS data, respectively. (For proportion of variance explained by the other unrotated principal components, see Supplementary Table S4.)

Figure 3 shows first unrotated principal component loadings of all psychopathological traits for the four genetic methods. The loadings on the first unrotated principal component mirrored the genetic correlations (Fig. 1): the average loadings were 0.75 for family data, 0.58 for GCTA, 0.57 for LDSC and 0.44 for GPS. We were able to test the statistical significance of loadings in family and GPS analyses, and found that all traits significantly loaded on the first unrotated principal component (all p-values $\leq 1.65 \times 10^{-41}$), even though the GPS data showed some of the lowest loadings. When we applied the conventional threshold of $\geq|0.30|$, we found that most of the loadings met this threshold: 100% of the disorders in family data, 80% in GCTA data, 88% in LDSC data, and 75% in GPS data. The variation in factor loadings across the four methods can be explained by the inclusion of different disorders, as average loadings for the disorders in common were highly similar (family = 0.70; GCTA = 0.69; LDSC = 0.66; GPS = 0.53).

Schizophrenia, Bipolar, and Depression consistently had the highest loadings on the first unrotated principal component across all genetic approaches with the exception of the GPS method, where Bipolar was not amongst the highest loading disorders.

Sensitivity analyses using LDSC and GPS data

To test whether GPS results changed when applying a different prior as part of the GPS calculation, we re-ran PCA using GPS based on the fraction of causal markers of 0.10. Results were almost identical (see Supplementary Table S5).

Furthermore, it is possible that low GPS loadings were attributable to insufficient statistical power, rather than a lack of true effects. Therefore, we re-ran PCAs using LDSC and GPS data based on superceded GWA study summary statistics with smaller sample sizes, where possible (see Supplementary Table S6 for sample information). Although we found a slight reduction in the variance explained by the first principal component in LDSC data (34 vs 35%), the effect was more pronounced in the GPS data (19 vs 22%). Additionally, average GPS loadings on the first principal component decreased from 0.44 to 0.37, and only 50% of the disorder GPS met the loading threshold of $\geq|0.30|$. These analyses suggest that as GWA study sample sizes increase, the magnitude of factor loading effect sizes on a genetic p factor will approach those derived from family studies.

Factor rotation solutions

Based on the criteria described in the Methods section, we retained two principal components for rotation for family, GCTA and GPS data, and three principal components for LDSC data (for more details, see Supplementary Table S4). However, to improve comparability of the rotated factor solutions across the four genetic methods, we kept two principal components for the LDSC data. Results of the rotation of three components for LDSC data can be found in Supplementary Table S7.

Figure 4 lists the loadings for the first two rotated factors after performing oblique rotation. Rotated factor loadings for all methods (family, GCTA, LDSC, GPS) show that Schizophrenia and Bipolar Disorder consistently load highly onto the same factor, together with Depression in the family and GCTA data. This is expected from the higher genetic intercorrelations between these traits for all methods (Fig. 1). For the remaining psychiatric traits, results were less consistent when comparing family data to genomic data (GCTA, LDSC, GPS). In part, this reflects the traits included—most notably, a Drug Abuse/Crime factor emerged from the family data because, unlike the other datasets, Drug Abuse, Alcohol Abuse and Violent Crime were included and created the first rotated factor. Anxiety also contributed to both

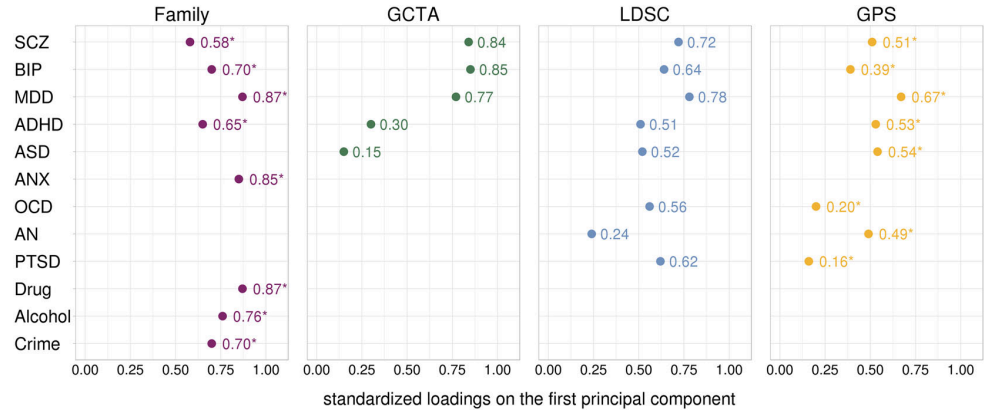


Fig. 3 Loadings of psychopathology traits on the first unrotated principal component for each of the four types of genetic data. *GCTA* Genome-wide Complex Trait Analysis, *LDSC* Linkage-Disequilibrium Score Regression, *GPS* Genome-wide Polygenic Score, *SCZ* Schizophrenia, *BIP* Bipolar Disorder, *MDD* Major Depressive Disorder, *ASD* Autism Spectrum Disorder, *ADHD* Attention-Deficit/Hyperactivity Disorder, *ANX* Anxiety, *OCD* Obsessive-Compulsive Disorder, *AN* Anorexia Nervosa, *PTSD* Post-Traumatic Stress Disorder; Drug = Drug Abuse; Alcohol = Alcohol Abuse; Crime = Convictions of Violent Crimes. * = reached statistical significance of $p \leq 1.65 \times 10^{-41}$; it was only possible to test the statistical significance for the loadings relating to GPS and family data (see Methods section for details)

rotated factors. For the LDSC and GPS method, which are based on the most powerful GWA studies, the second factor primarily included Depression, Attention-Deficit/Hyperactivity Disorder, Autism and Post-Traumatic Stress Disorder. Correlations between the first and second oblique rotated factors were 0.45 for family data, 0.08 for GCTA data, 0.14 for LDSC data and 0.10 for GPS data.

Discussion

These results provide genetic support for *p*, a general factor of psychopathology that represents a single, continuous genetic dimension of the psychiatric spectrum. The four methods used to estimate genetic correlations differ substantially: quantitative genetic analysis of siblings and half-siblings[12], GCTA estimates based on SNP differences between unrelated individuals[14], LDSC analysis based on GWA summary statistics, and GPS for individual data presented in this paper. Nonetheless, each of the principal component analyses from the four methods yielded a general factor on which all disorders loaded, explaining between 20 and 60% of the total variance.

Schizophrenia, Bipolar and Depression are the oldest and most consistently diagnosed psychiatric disorders, yet they are consistently among the highest- loading disorders on this genetic *p* factor. This finding is unlikely to be due to some artifact of genetic analysis because it was consistent across different genetic methods applied to different samples.

It is difficult to draw general conclusions about the other disorders that varied across the four genetic methods (Obsessive Compulsive Disorder, Anorexia, and Post-Traumatic Stress Disorder, Anxiety, Drug Abuse, Alcohol Abuse and Violent Crime). However, when any of these disorders were included in a study, they consistently contributed to a genetic *p* factor in the sense that they loaded positively on the first unrotated principal component.

Although the four genetic methods yielded similar patterns of correlations and patterns of loadings on the first unrotated principal component, they differed in the magnitude of their estimates of correlations and loadings, even when only considering the disorders in common (i.e., Schizophrenia, Bipolar, Depression, Autistic Spectrum Disorder). In principle, genetic correlations calculated through GCTA and LDSC should not differ substantially from family study estimates. Even though univariate SNP-h^2 is generally lower than family-h^2 because the SNP-h^2 estimate does not include rare variants and nonadditive effects, this downward bias influences both numerator and denominator to equal extents when calculating genetic correlations ($r_g = h_x h_y / \sqrt{h_{x^2} h_{y^2}}$), therefore cancelling out the bias[35]. However, if the correlation between causal SNPs is stronger for common variants than for rare variants, the SNP genetic correlation estimate would be higher than family study estimates, because only common SNPs are included in the analysis[16]. Nevertheless, for the disorders in common, family data produced higher average genetic correlations (0.49) than GCTA (0.34) and LDSC (0.37). An alternative explanation involves differences in sample ascertainment and psychiatric diagnoses. In most genomic studies, sampling strategies may select 'pure' cases and exclude cases with other co-occurring conditions, and such 'pure' cases do not represent the disordered population[36]. In contrast, family data used in

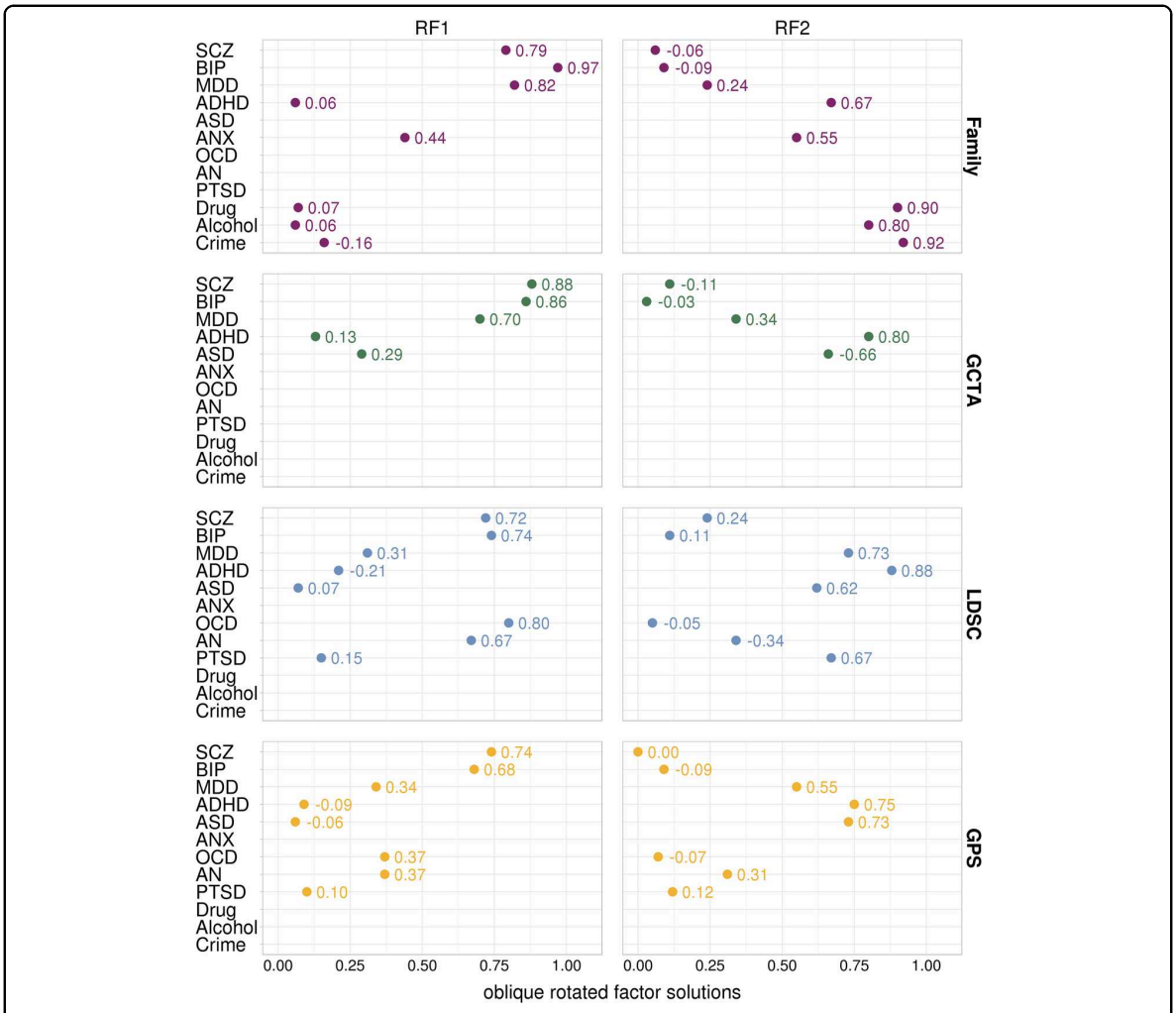


Fig. 4 Rotated factor loadings for the four types of genetic data. *RF* rotated factor based on oblique (*Oblimin*) rotation, *GCTA* Genome-wide Complex Trait Analysis, *LDSC* Linkage-Disequilibrium Score Regression, *GPS* Genome-wide Polygenic Score, *SCZ* Schizophrenia, *BIP* Bipolar Disorder, *MDD* Major Depressive Disorder, *ASD* Autism Spectrum Disorder, *ADHD* Attention-Deficit/Hyperactivity Disorder, *ANX* Anxiety, *OCD* Obsessive-Compulsive Disorder, *AN* Anorexia Nervosa, *PTSD* Post-Traumatic Stress Disorder; Drug = Drug Abuse; Alcohol = Alcohol Abuse; Crime = Convictions of Violent Crimes

this study[12] were based on a non-hierarchical approach to classification, thus allowing for greater overlap among the disorders.

GPS results, which are based on the most conceptually distinct method, yielded the lowest overall correlations. A GPS is the aggregation of all genetic effects found in an independent GWA analysis in respect to an individual's genotype. Therefore, GPS correlations index the extent to which the total variance of individuals' GPS for one trait covaries with GPS for other traits. Two possible reasons why GPS correlations may be the lowest are that (i) in addition to true effects, a GPS includes the measurement error for all the SNPs tested across the genome in GWA analysis and (ii) a GPS is generated using genotypes from one cohort and effect sizes from a second, independent cohort.

What causes this genetic *p* factor? The positive manifold of the genetic *p* factor is agnostic about its causes. There are several, equally plausible hypotheses for the mechanisms that cause cross-disorder correlations[37]. One possible pathway may be *biological pleiotropy*, where DNA variants are causally involved in the development of several traits related to psychopathology. An alternative explanation is *mediated pleiotropy*, in which comorbidity occurs because DNA variants increase risk for one disorder, and then this disorder causes other disorders in

turn. A third hypothesis is that DNA variants cause some general impairment that forms the core of various disorders, consequently producing genetic correlation between specific diagnoses. That is, the thousands of DNA variants associated with each symptom or disorder might affect all personality and cognitive processes that increase risk, thus providing many pathways to psychopathology.

Although it is remarkable how much genetic variance is explained by *p*, it does not explain all, or even most, of the genetic variance. Assuming a hierarchical model with *p* at the highest level[6,7], broader psychiatric dimensions at a middle level, and specific psychopathologies at the lowest level, the question is how much genetic variance is accounted for by the three levels. In the realm of cognitive abilities, there continues to be debates about the nature of the middle level[38].

As compared to *p*, there is less clarity in our results about the nature of the second level of the hierarchical structure, as represented by the rotated factor solutions. One rotated factor consistently includes Schizophrenia and Bipolar Disorder. However, the other rotated factor is less clear. For example, although Attention-Deficit/Hyperactivity Disorder loads on the second factor, it clusters positively with Depression and Autism Spectrum Disorder in the LDSC and GPS results, positively with Anxiety, substance abuse and Crime in the family results, and negatively with Autistic Spectrum Disorder in the GCTA and GPS results. It may be that the second level of the hierarchical structure will remain unclear until analyses of this type begin to use a transdiagnostic approach, that is, using symptoms to build a hierarchical model from the ground up. As these data become available in the future, we will be able test the genetic *p* factor model more formally by contrasting it to alternative models.

Another issue for future research is the extent to which the *p* factor is even more general than psychiatric disorders. The same approach can be used to investigate the genetic relationship between psychiatric disorders and personality traits, cognitive traits, structural and functional brain traits, medical and neurological disorders, and physiological traits. However, here we chose to focus on the extent to which a genetic *p* factor emerges from genomic analyses of psychiatric disorders themselves.

As noted, our analyses are limited to the data that currently exist, including the power of current GWA studies and the disorders included in these studies. A fundamental limitation is 'missing heritability', the gap between SNP-h^2 and family study heritability estimates. We used the most recent publicly available GWA summary statistics, some of which are considerably underpowered. This limitation most affects our GPS analysis, which predicts genetic risk at the level of individuals. The modest SNP-h^2 and measurement error of the GWA studies from which the GPS were derived are partly responsible for the low correlations between the GPS. More powerful GWA studies are in progress, and we are optimistic that new GPS will have improved predictive accuracy. More generally, GWA studies focused on phenotypic *p* should be able to capture genetic *p* to a greater extent than trying to derive genetic *p* from GWA studies of separate disorders that are sometimes diagnosed as 'pure' cases that exclude other diagnoses.

In conclusion, we report strong evidence for a genetic *p* factor that represents a continuous, underlying dimension of psychiatric risk using four distinct genetic methods. As GWA studies continue to increase in sample size as well as in the diversity of their target traits, our current results suggest that a genetic *p* factor could be useful in psychiatric research.

Acknowledgements

We gratefully acknowledge the ongoing contribution of the participants in the Twins Early Development Study (TEDS) and their families. The authors also gratefully acknowledge the contribution of Erik Pettersson who provided genetic correlations from his Swedish family study. TEDS is supported by a program grant to RP from the UK Medical Research Council (MR/M021475/1 and previously G0901245), with additional support from the US National Institutes of Health (AG046938). The research leading to these results has also received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/ grant agreement 602768 and ERC grant agreement 295366. SS is supported by the MRC/IoPPN Excellence Award and by the US National Institutes of Health (AG046938). This study represents independent research part funded by the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health. High performance computing facilities were funded with capital equipment grants from the GSTT Charity (TR130505) and Maudsley Charity (980).

Author details

[1]MRC Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK. [2]NIHR Biomedical Research Centre for Mental Health, South London and Maudsley NHS Trust, London, UK. [3]Department of Psychology and Neuroscience, Duke University, Durham, USA. [4]Center for Genomic and Computational Biology, Duke University, Durham, USA. [5]Department of Psychiatry and Behavioral Sciences, Duke University Medical Center, Durham, USA

Author contributions

Study concept and design: S.S., R.P. Processed and quality controlled genotype data: S.S. Analysis of data: S.S. Interpretation of data: All authors. Wrote the paper: S.S., R.P. Contributed to and critically reviewed the manuscript: All authors.

Conflict of interest

The authors declare that they have no conflict of interest.

Supplementary Information accompanies this paper at (https://doi.org/10.1038/s41398-018-0217-4).

Received: 21 June 2018 Accepted: 16 July 2018
Published online: 02 October 2018

References

1. Kessler, R. C., Chiu, W. T., Demler, O. & Walters, E. E. Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Arch. Gen. Psychiatry* **62**, 617–627 (2005).
2. Kotov, R. et al. The hierarchical taxonomy of psychopathology (HiTOP): a dimensional alternative to traditional nosologies. *J. Abnorm. Psychol.* **126**, 454–477 (2017).
3. Wright, A. G. C. et al. The structure of psychopathology: toward an expanded quantitative empirical model. *J. Abnorm. Psychol.* **122**, 281–294 (2013).
4. Lahey, B. B. et al. Is there a general factor of prevalent psychopathology during adulthood? *J. Abnorm. Psychol.* **121**, 971–977 (2012).
5. Caspi, A. et al. The *p* factor: one general psychopathology factor in the structure of psychiatric disorders? *Clin. Psychol. Sci.* **2**, 119–137 (2014).
6. Caspi, A., Moffitt, T. E. All for one and one for all: Mental disorders in one dimension. *Am. J. Psychiatry*. https://doi.org/10.1176/appi.ajp.2018.17121383 (2018).
7. Lahey, B. B., Krueger, R. F., Rathouz, P. J., Waldman, I. D. & Zald, D. H. A hierarchical causal taxonomy of psychopathology across the life span. *Psychol. Bull.* **143**, 142–186 (2017).
8. McLaughlin, K. A. et al. Parent psychopathology and offspring mental disorders: results from the WHO World Mental Health Surveys. *Br. J. Psychiatry* **200**, 290–299 (2012).
9. Kendler, K. S. Major depression and generalised anxiety disorder—same genes, (partly) different environments—revisited. *Br. J. Psychiatry* **168**, 68–75 (1996).
10. Lichtenstein, P. et al. Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. *Lancet* **373**, 234–239 (2009).
11. Kendler, K. S. et al. The structure of genetic and environmental risk factors for syndromal and subsyndromal common DSM-IV axis I and all axis II disorders. *Am. J. Psychiatry* **168**, 29–39 (2011).
12. Pettersson, E., Larsson, H., & Lichtenstein, P. Common psychiatric disorders share the same genetic origin: a multivariate sibling study of the Swedish population. *Mol. Psychiatry* **21**, 717–721 (2016).
13. Cross-Disorder, Group of the Psychiatric Genomics Consortium. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet* **381**, 1371–1379 (2013).
14. Lee, S. H. et al. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat. Genet.* **45**, 984–994 (2013).
15. Bulik-Sullivan, B. K. et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
16. Bulik-Sullivan, B. et al. An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
17. Anttila, V. et al. Analysis of shared heritability in common disorders of the brain. *Science*. https://doi.org/10.1126/science.aap8757 (2018).
18. Krapohl, E. et al. Phenome-wide analysis of genome-wide polygenic scores. *Mol. Psychiatry* **21**, 1188–1193 (2016).
19. Dudbridge, F. Polygenic epidemiology. *Genet. Epidemiol.* **40**, 268–272 (2016).
20. Haworth, C. M. A., Davis, O. S. P. & Plomin, R. Twins Early Development Study (TEDS): a genetically sensitive investigation of cognitive and behavioral development from childhood to young adulthood. *Twin. Res. Hum. Genet.* **16**, 117–125 (2013).
21. Selzam, S. et al. Predicting educational achievement from DNA. *Mol. Psychiatry* **22**, 267–272 (2017).
22. Vilhjalmsson, B. J. et al. Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *Am. J. Hum. Genet.* **97**, 576–592 (2015).
23. Lee, S. H., Yang, J., Goddard, M. E., Visscher, P. M. & Wray, N. R. Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. *Bioinformatics* **28**, 2540–2542 (2012).
24. Visscher, P. M. et al. Statistical power to detect genetic (co)variance of complex traits using SNP data in unrelated samples. *PLOS Genet.* https://doi.org/10.1371/journal.pgen.1004269 (2014).
25. Jolliffe, I. T. Principal Component Analysis and Factor Analysis. In Principal Component Analysis. *Springer Series in Statistics*. Springer: New York, NY, pp. 115–128 (1986).
26. Yamamoto, H. et al. Statistical hypothesis testing of factor loading in principal component analysis and its application to metabolite set enrichment analysis. *BMC Bioinform.* https://doi.org/10.1186/1471-2105-15-51 (2014).
27. Yong, A. G. & Pearce, S. A beginner's guide to factor analysis: focusing on exploratory factor analysis. *Tutor. Quant. Methods Psychol.* **9**, 79–94 (2013).
28. Kaiser, H. F. The application of electronic computers to factor analysis. *Educ. Psychol. Meas.* **20**, 141–151 (1960).
29. Horn, J. L. A rationale and test for the number of factors in factor analysis. *Psychometrika* **30**, 179–185 (1965).
30. Cattell, R. B. The scree test for the number of factors. *Multivar. Behav. Res.* **1**, 245–276 (1966).
31. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.r-project.org/ (2017).
32. Huang, F. hornpa: Horn's (1965) test to determine the number of components/factors. https://CRAN.R-project.org/package=hornpa (2015).
33. Revelle, W. R. psych: procedures for personality and psychological research. https://CRAN.R-project.org/package=psych (2017).
34. Bernaards, C. A. & Jennrich, R. I. Gradient projection algorithms and software for arbitrary rotation criteria in factor analysis. *Educ. Psychol. Meas.* **65**, 676–696 (2005).
35. Trzaskowski, M. et al. DNA evidence for strong genome-wide pleiotropy of cognitive and learning abilities. *Behav. Genet.* **43**, 267–273 (2013).
36. Newman, D. L., Moffitt, T. E., Caspi, A. & Silva, P. A. Comorbid mental disorders: implications for treatment and sample selection. *J. Abnorm. Psychol.* **107**, 305–311 (1998).
37. Solovieff, N., Cotsapas, C., Lee, P. H., Purcell, S. M. & Smoller, J. W. Pleiotropy in complex traits: challenges and strategies. *Nat. Rev. Genet.* **14**, 483–495 (2013).
38. Johnson, W. & Bouchard, T. The structure of human intelligence: It is verbal, perceptual, and image rotation (VPR), not fluid and crystallized. *Intelligence* **33**, 393–416 (2005).

Chapter 5 – Evidence for gene-environment correlation in child feeding: Links between common genetic variation for BMI in children and parental feeding practices

This chapter is presented as a published paper. It is an exact copy of this publication:

Selzam, S., McAdams, T. A., Coleman, J. R., Carnell, S., O'Reilly, P. F., Plomin, R., & Llewellyn, C. H. (2018). Evidence for gene-environment correlation in child feeding: Links between common genetic variation for BMI in children and parental feeding practices. *PLoS genetics*, 14(11), e1007757. doi: 10.1371/journal.pgen.1007757

Supplemental materials are included under Appendix 4.

PLOS GENETICS

RESEARCH ARTICLE

Evidence for gene-environment correlation in child feeding: Links between common genetic variation for BMI in children and parental feeding practices

Saskia Selzam[1]*, Tom A. McAdams[1], Jonathan R. I. Coleman[1,2], Susan Carnell[3], Paul F. O'Reilly[1], Robert Plomin[1], Clare H. Llewellyn[4]*

1 Institute of Psychiatry, Psychology and Neuroscience, MRC Social, Genetic and Developmental Psychiatry Centre, King's College London, London, United Kingdom, **2** NIHR Biomedical Research Centre for Mental Health, South London and Maudsley NHS Trust, London, United Kingdom, **3** Department of Psychiatry and Behavioral Sciences, Division of Child and Adolescent Psychiatry, Johns Hopkins University School of Medicine, Baltimore, Maryland, United States of America, **4** Department of Behavioural Science and Health, University College London, London, United Kingdom

* saskia.selzam@kcl.ac.uk (SS); c.llewellyn@ucl.ac.uk(CHL)

OPEN ACCESS

Citation: Selzam S, McAdams TA, Coleman JRI, Carnell S, O'Reilly PF, Plomin R, et al. (2018) Evidence for gene-environment correlation in child feeding: Links between common genetic variation for BMI in children and parental feeding practices. PLoS Genet 14(11): e1007757. https://doi.org/10.1371/journal.pgen.1007757

Editor: Heather J. Cordell, Newcastle University, UNITED KINGDOM

Received: April 17, 2018

Accepted: October 11, 2018

Published: November 20, 2018

Data Availability Statement: The dataset analysed during the current study is restricted due to participant confidentially. Restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. The Twins Early Development Study is a third party that owns the data, and information about data access is available here: https://www.teds.ac.uk/research/collaborators-and-data/teds-data-access-policy. Requests should be made via the data request form supplied, which

Abstract

The parental feeding practices (PFPs) of excessive restriction of food intake ('restriction') and pressure to increase food consumption ('pressure') have been argued to causally influence child weight in opposite directions (high restriction causing overweight; high pressure causing underweight). However child weight could also 'elicit' PFPs. A novel approach is to investigate gene-environment correlation between child genetic influences on BMI and PFPs. Genome-wide polygenic scores (GPS) combining BMI-associated variants were created for 10,346 children (including 3,320 DZ twin pairs) from the Twins Early Development Study using results from an independent genome-wide association study meta-analysis. Parental 'restriction' and 'pressure' were assessed using the Child Feeding Questionnaire. Child BMI standard deviation scores (BMI-SDS) were calculated from children's height and weight at age 10. Linear regression and fixed family effect models were used to test between- (n = 4,445 individuals) and within-family (n = 2,164 DZ pairs) associations between the GPS and PFPs. In addition, we performed multivariate twin analyses (n = 4,375 twin pairs) to estimate the heritabilities of PFPs and the genetic correlations between BMI-SDS and PFPs. The GPS was correlated with BMI-SDS ($\beta = 0.20$, $p = 2.41\text{x}10^{-38}$). Consistent with the gene-environment correlation hypothesis, child BMI GPS was positively associated with 'restriction' ($\beta = 0.05$, $p = 4.19\text{x}10^{-4}$), and negatively associated with 'pressure' ($\beta = -0.08$, $p = 2.70\text{x}10^{-7}$). These results remained consistent after controlling for parental BMI, and after controlling for overall family contributions (within-family analyses). Heritabilities for 'restriction' (43% [40–47%]) and 'pressure' (54% [50–59%]) were moderate-to-high. Twin-based genetic correlations were moderate and positive between BMI-SDS and 'restriction' ($r_A = 0.28$ [0.23–0.32]), and substantial and negative between BMI-SDS and 'pressure' ($r_A = -0.48$ [-0.52 - -0.44]. Results suggest that the degree to which parents limit or encourage children's food intake is partly influenced by children's genetic predispositions

are then reviewed by the TEDS Executive committee. Data will be made available on request to interested researchers in the same fashion in which it was made available to the authors, allowing data sharing for novel collaborations leading to new publications.

Funding: We gratefully acknowledge the ongoing contribution of the participants in the Twins Early Development Study (TEDS) and their families. TEDS is supported by a program grant to RP from the UK Medical Research Council (MR/M021475/1 and previously G0901245), with additional support from the US National Institutes of Health (AG046938). The research leading to these results has also received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/ grant agreement n° 602768 and ERC grant agreement n° 295366. RP is supported by a Medical Research Council Professorship award (G19/2). SS is supported by the MRC/IoPPN Excellence Award and by the US National Institutes of Health (AG046938). High performance computing facilities were funded with capital equipment grants from the GSTT Charity (TR130505) and Maudsley Charity (980). TAM is supported by a Sir Henry Dale Fellowship, jointly funded by the Wellcome Trust and the Royal Society (107706/Z/15/Z). This study represents independent research part-funded by the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health. High performance computing facilities were funded with capital equipment grants from the GSTT Charity (TR130505) and Maudsley Charity (980). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist

to higher or lower BMI. These findings point to an evocative gene-environment correlation in which heritable characteristics in the child elicit parental feeding behaviour.

Author summary

It is widely believed that parents influence their child's BMI via certain feeding practices. For example, rigid restriction has been argued to cause overweight, and pressuring to eat to cause underweight. However, recent longitudinal research has not supported this model. An alternative hypothesis is that child BMI, which has a strong genetic basis, evokes parental feeding practices ('gene-environment correlation'). To test this, we applied two genetic methods in a large sample of 10-year-old children from the Twins Early Development Study: a polygenic score analysis (DNA-based score of common genetic variants associated with BMI in genome-wide meta-analyses), and a twin analysis (comparing resemblance between identical and non-identical twin pairs). Polygenic scores correlated positively with parental restriction of food intake ('restriction'; $\beta = 0.05$, $p = 4.19\text{x}10^{-4}$), and negatively with parental pressure to increase food intake ('pressure'; $\beta = -0.08$, $p = 2.70\text{x}10^{-7}$). Associations were unchanged after controlling for all genetic and environmental effects shared within families. Results from twin analyses were consistent. 'Restriction' (43%) and 'pressure' (54%) were substantially heritable, and a positive genetic correlation between child BMI and 'restriction' ($r_A = 0.28$), and negative genetic correlation between child BMI and 'pressure' ($r_A = -0.48$) emerged. These findings challenge the prevailing view that parental behaviours are the sole cause of child BMI by supporting an alternate hypothesis that child BMI also causes parental feeding behaviour.

Introduction

The home and family environment has been studied for decades with the assumption that it is a crucial determinant of children's health and development. Since the onset of the childhood obesity crisis at the turn of the century, the spotlight has turned onto environmental factors associated with variation in adiposity, in the hope that modifiable elements may be identified as intervention targets. Perhaps unsurprisingly, parental behaviours have received a great deal of attention. Parents are widely considered to be the 'gatekeepers' to their children's food, and powerful shapers of their developing eating behaviour[1–3]. Two types of parental feeding practices (PFPs) in particular have been hypothesised to play a causal role in children's ability to develop good self-regulation of food intake and consequently determine their weight. Excessive restriction of the type and amount of food a child is allowed to eat ('restriction') has been hypothesised to lead to overeating when parental restriction is no longer in place, because the child will potentially then hanker after the foods he or she is not usually allowed to eat–the so-called 'forbidden fruit effect'[1,4,5]. On the other hand, overly pressuring a child to eat, or to finish everything on the plate ('pressure'), is thought to be anxiety-provoking for a child with a poor appetite, and serves only to increase undereating further, and compromise weight gain [6,7].

A wealth of cross-sectional findings are consistent with these hypotheses[8], but another plausible explanation for the observed correlations is that parents are responding to their child's emerging characteristics, not simply causing them. Parents may only adopt restrictive strategies when a child shows a tendency toward overeating, or gains excessive weight; and

they may pressure their child to eat only if he or she is a poor eater, or has underweight. The few longitudinal studies testing bidirectionality have shown that children's weight prospectively predicts PFPs[9–13]. Furthermore, three studies showed no prospective association from PFPs to child weight[10], and the studies reporting bidirectional relationships found stronger associations from child weight to parental behaviour than the reverse direction[9,11]. Although these findings point towards children's weight eliciting PFPs, the possibility of residual confounding in observational studies hinders conclusions about causation–temporality does not necessarily mean causality.

Testing whether children genuinely cause their parents' behaviour presents challenges. It is not possible–practically or ethically–to randomise children to have overweight or underweight, and examine how parents respond. Genetic approaches provide a powerful alternative method of interrogating the role of children in causing their parents' behaviour towards them, especially for child characteristics with an established genetic basis. To date, no study has applied genetically sensitive methods to test for gene-environment correlation in parental feeding behaviour. Family and twin studies have shown that Body Mass Index (BMI), is highly heritable in both adulthood and late childhood (~70%)[14–16]. Twin designs can also be used to test if parental behaviour has a heritable component, by comparing within-pair resemblance for identical and fraternal twin pairs in childhood. If found, this indicates that parental behaviour is explained to some extent by variation in children's genotype–termed evocative gene-environment correlation[17]. Twin designs can also be extended to the analysis of multiple variables to establish if genetic influence on a particular child characteristic (e.g. weight) also predicts the parental behaviour of interest (e.g. PFPs). If such analyses show that a child characteristic is genetically correlated with parenting traits, it indicates that these child characteristics influence parenting behaviours. A meta-analysis of 32 twin studies of different types of parenting behaviour reported an average heritability estimate of 24%, indicating that children's genotype is predictive of a moderate amount of variation in parental behaviour[18].

Children's DNA can also be used to test for gene-environment correlation. Genome-wide meta-analyses have made great progress in identifying common single nucleotide polymorphisms (SNPs) that are associated with body mass index (BMI) in adults and children[19]. These can be combined to calculate a genome-wide polygenic score (GPS) that indexes individual-specific propensity to higher or lower BMI, along a continuum, although in the aggregate the GPS explains only a small proportion of variance in BMI (approximately 3%)[20]. Nevertheless, children's BMI GPS can therefore be used to test the hypothesis that parents develop their feeding practices specifically in response to their child's weight, as indicated by a correlation between child BMI GPS and PFPs. A caveat to this is that a parent's feeding practices may reflect their own genetic predisposition to be of a higher or lower BMI, rather than that of their children. In this way, a correlation between child BMI GPS and PFPs may simply reflect a child's genetic predisposition to be of a higher or lower BMI, which they inherit from their parent with whom they share 50% of their DNA. In addition, genetic effects related to adult BMI discovered in genome-wide association studies could potentially incorporate effects of PFPs if they were to causally influence child BMI, and its trajectory into adulthood. However, within-family designs can circumvent both of these limitations to some extent. Studying variation in PFPs according to variation in BMI GPS within non-identical co-twins accounts for both genetic and environmental shared effects within families (e.g. parental genetic predisposition to be of higher or lower BMI). By applying both quantitative and molecular genetic methods, and utilising statistical approaches to account for shared family effects, we intended to address the various limitations presented by the individual methods.

The goals of this study were to test for gene-environment correlation between children's BMI and PFPs, using a twin design and a BMI GPS. We hypothesised that: (i) children's BMI

GPS would be positively associated with parental restriction and negatively associated with parental pressure, even after accounting for shared genetic and environmental family influences; and (ii) parental restriction and parental pressure would be moderately heritable, and that genetic influence on PFPs would be partly explained by genetic influence on children's BMI.

Results

Phenotypic correlations

Child BMI-SDS was significantly positively correlated with 'restriction' ($\beta = 0.19$, $t(4004) = 12.09$, $p = 4.45\text{x}10^{-33}$, $R^2 = 0.035$), such that parents were more restrictive over their child's food intake if the child had a higher BMI. In contrast, child BMI-SDS was significantly negatively correlated with 'pressure' ($\beta = -0.24$, $t(4058) = -15.59$, $p = 3.14\text{x}10^{-53}$, $R^2 = 0.056$), such that parents exerted higher amounts of pressure on their child to eat, if their child was leaner. 'Restriction' and 'pressure' were significantly positively correlated ($\beta = 0.15$, $t(4207) = 9.51$, $p = 3.08\text{x}10^{-21}$, $R^2 = 0.021$), suggesting that parents who tend to exert higher levels of 'restriction' also exert a more pressuring feeding style, to some extent.

Genome-wide polygenic score (GPS) analyses

In our sample of unrelated individuals, child BMI GPS was positively correlated with child BMI-SDS ($\beta = 0.20$, $t(4226) = 13.08$, $p = 2.41\text{x}10^{-38}$, $R^2 = 0.039$). Mirroring phenotypic results for child BMI-SDS, children's BMI GPS was significantly positively correlated with 'restriction' ($\beta = 0.05$, $t(4255) = 3.53$, $p = 4.19\text{x}10^{-4}$, $R^2 = 0.003$), and significantly negatively correlated with 'pressure' ($\beta = -0.08$, $t(4315) = -5.15$, $p = 2.70\text{x}10^{-7}$, $R^2 = 0.006$) (Fig 1). These findings indicate

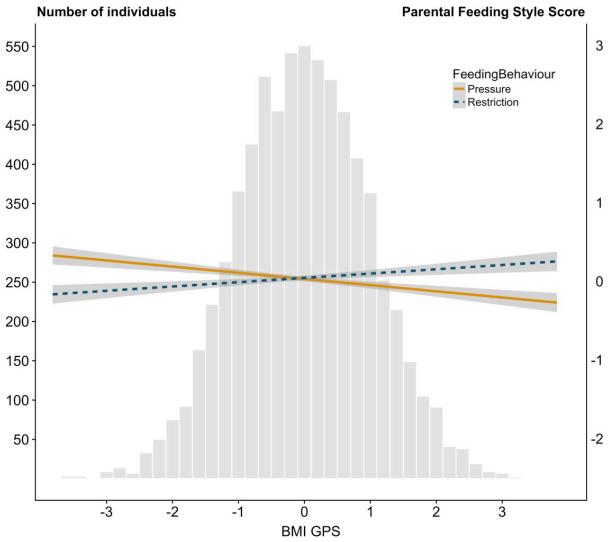


Fig 1. The associations between child BMI polygenic score and parental feeding practices. Child BMI GPS predicting standardized measures of parental 'restriction' ($\beta = 0.05$, $p = 4.19\text{x}10^{-4}$) and parental 'pressure' ($\beta = -0.08$, $p = 2.70\text{x}10^{-7}$) as indicated by the best-fit regression lines. The grey areas surrounding the best-fit lines represent standard errors of the prediction estimates. The histogram depicts the BMI GPS normal distribution.

https://doi.org/10.1371/journal.pgen.1007757.g001

that children's genetic predisposition to higher BMI, elicits, to some extent, restrictive feeding behaviours in the parent; whereas children's genetic predisposition to lower BMI elicits greater pressure to eat by parents.

Parental BMI correlated positively with child BMI-SDS ($\beta = 0.26$, $t(3761) = 17.00$, $p = 1.57\text{x}10^{-62}$, $R^2 = 0.071$) and 'restriction' ($\beta = 0.08$, $t(3711) = 4.64$, $p = 3.65\text{x}10^{-6}$, $R^2 = 0.005$), but was not significantly associated with 'pressure' ($\beta = -0.03$, $t(3757) = -1.68$, $p = 0.09$, $R^2 < 0.001$). The magnitude and direction of effects remained identical after controlling for parental BMI in 'restriction' ($\beta = 0.05$, $t(3711) = 2.92$, $p = 3.48\text{x}10^{-3}$, $R^2 = 0.003$) and in 'pressure' ($\beta = -0.08$, $t(3757) = -4.62$, $p = 3.97\text{x}10^{-6}$, $R^2 = 0.005$).

Within-family analysis

To establish the association between children's BMI GPS and PFPs entirely without confounding by genetic and environmental family factors shared by twin pairs, we performed family fixed-effect analyses in dizygotic (DZ) co-twins. This analysis examined the extent to which parents vary their 'restriction' and 'pressure' across twin pairs in response to differences in their BMI GPS. As shown in Fig 2, beta coefficients for BMI GPS predicting PFPs remained largely stable when comparing unrelated individuals (Model 1) and DZ twin pairs (Model 2). For unrelated individuals (Model 1) child BMI-SDS significantly positively predicted 'restriction' and significantly negatively predicted 'pressure', as previously reported. The magnitude of the within-family estimates for the combined (same-sex and opposite-sex) DZ co-twins (Model 2) were virtually the same as those for the unrelated individuals for the relationships between BMI GPS and 'restriction' ($t(2054) = 3.50$, $p = 7.10\text{x}10^{-3}$, $Adj.\ R^2_{model} = 0.724$) and BMI GPS and 'pressure' ($t(2103) = -4.82$, $p = 1.52\text{x}10^{-6}$, $Adj.\ R^2_{model} = 0.641$) ($R^2$ magnitudes for Model 2 are large because all shared factors among family members, including genetic and environmental influences, are accounted for). These findings indicate that even when shared family effects are completely accounted for, children's BMI GPS is significantly associated with PFPs, providing additional evidence that children's genetic predisposition to BMI evokes certain parental feeding responses. When repeating Model 2 analyses separately for same-sex and opposite-sex DZs, magnitudes of effect sizes (Fig 2) remained consistent for the prediction of 'pressure' in same-sex DZ pairs ($t(1118) = -3.36$, $p = 8.02\text{x}10^{-4}$, $Adj.\ R^2_{model} = 0.607$) and opposite-sex DZ pairs ($t(984) = -3.49$, $p = 5.12\text{x}10^{-4}$, $Adj.\ R^2_{model} = 0.678$). Although BMI GPS in opposite-sex DZs was a significant predictor of within-family differences in 'restriction' ($t(966) = 3.76$, $p = 1.82\text{x}10^{-4}$, $Adj.\ R^2_{model} = 0.731$), same-sex DZ data did not show a significant within-family association ($t(1087) = 1.21$, $p = 0.23$, $Adj.\ R^2_{model} = 0.719$), indicating that within a family environment, GPS differences in BMI between same-sex DZ twins are not related to differences in parental 'restriction'.

Twin analysis

We performed multivariate genetic analyses (a correlated factors model) to establish the heritability of 'restriction' and 'pressure' and to test the extent to which genetic influence on child BMI-SDS elicited PFPs as indicated by the magnitude of genetic correlations between BMI, 'restriction', and 'pressure'. Fig 3 shows the variance components (A, C and E) for each measured phenotype, as well as the genetic, shared environmental and non-shared environmental correlations between phenotypes derived from the correlated factors model (see Supplementary S1 Table for fit statistics and model comparisons, and Supplementary S2 Table for intraclass correlations). Heritability estimates (A) were moderate to high for parental 'restriction' (43%, 95% CI [40%, 47%]) and parental 'pressure' (54%, 95% CI [50%, 59%]); heritability of child BMI-SDS was high (78%, 95% CI [72%, 84%]). Consistent with the findings from the

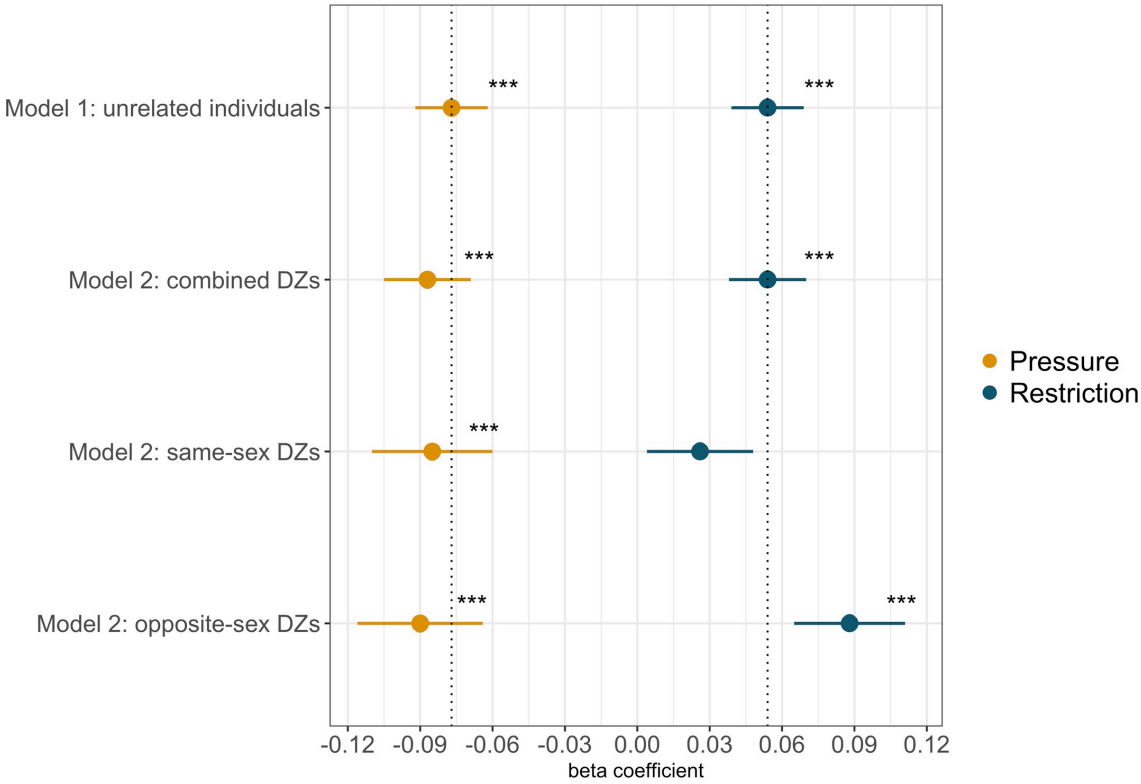


Fig 2. Contrasting results from between-family analyses to results from within-family analyses. Model 1 describes results using BMI GPS of unrelated individuals to predict PFPs, where β_{GPS} indicates the change in the outcome trait per one standard deviation increase in the BMI GPS. Model 2 summarises results using BMI genome-wide polygenic scores in a sample of DZ co-twins using a family fixed-effects model, where β_{GPS} indicates the increase in PFPs within DZ pairs, per one standard deviation increase in BMI GPS within DZ pairs. Model 2 analyses were performed using the combined DZ sample, and same-sex DZ pairs and opposite-sex DZ pairs only. The dotted lines represent the beta coefficient estimates for Model 1. * = $p<0.05$; ** = $p<0.01$; *** = $p<0.001$.

https://doi.org/10.1371/journal.pgen.1007757.g002

GPS analyses, there was a significant, positive moderately sized genetic correlation between child BMI-SDS and parental 'restriction' (r_A = 0.28, 95% CI [0.23, 0.32]), indicating that some of the genetic effects that predispose a child to a higher BMI also elicit more food restriction by their parent. A sizeable significant negative genetic correlation was observed between child BMI-SDS and parental 'pressure' (r_A = -0.48, 95% CI [-0.52, -0.44]), indicating that many of the genetic effects that predispose a child to a lower BMI elicit greater parental pressure on the child to eat.

Monozygotic (MZ) twin discordance analysis

As shown in the twin analyses (Fig 3 and Supplementary S3 Table), variation in child BMI-SDS is partly caused by non-shared environmental influences, which correlate significantly with non-shared environmental influences for 'restriction' (r_E = 0.20) and 'pressure' (r_E = -0.29). We therefore performed MZ twin difference analyses to examine these relationships more closely. In contrast to child BMI-SDS MZ difference scores, most twins did not differ in their PFP (Supplementary S1 Fig). Nevertheless, we found that child BMI-SDS difference scores

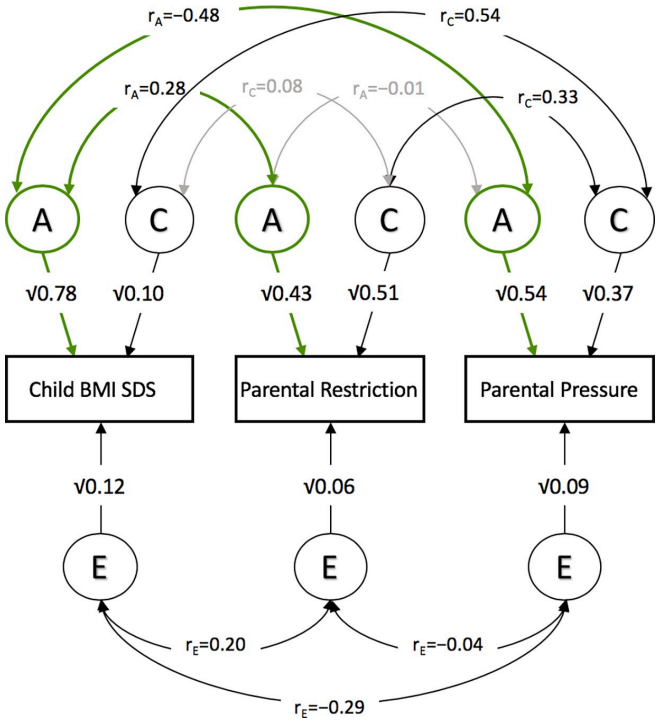


Fig 3. The correlated factors model. A correlated factors model (males and females combined) showing: (i) the genetic (A), shared environmental (C) and non-shared environmental (E) influences on child BMI SDS, parental restriction and pressure; and (ii) common genetic (r_A), shared environmental (r_C) and non-shared environmental (r_E) correlations between child BMI, and parental restriction and pressure. Grey arrows indicate non-significant associations. Correlations including the 95% confidence intervals can be found in Supplementary S3 Table.

https://doi.org/10.1371/journal.pgen.1007757.g003

predicted both differences in 'restriction' (β = 0.14, $t(1484)$ = 7.98, p = 2.88x10^{-15}, R^2 = 0.041) and 'pressure' (β = -0.25, $t(1498)$ = -12.26, p = 5.12x10^{-33}, R^2 = 0.09). These findings suggest that there are common non-shared environmental sources of variance for both PFP and child BMI; within identical twin pairs who share 100% of their genetic and shared environmental influence, parents apply more restrictive feeding practices on the twin with the higher BMI, and more pressuring feeding practices on the twin with the lower BMI score.

Discussion

Summary of findings

We describe the first study to test for gene-environment correlation for parental feeding behaviour in relation to child weight, using a twin design and children's DNA. Results support our hypothesis that parents' feeding practices are evoked, in part, by their children. Parental 'restriction' and 'pressure' were positively and negatively associated with child BMI respectively, in keeping with many previous cross-sectional studies[8]. We applied novel genetic methods to show, for the first time, that children's BMI GPS was significantly positively associated with 'restriction' and negatively associated with 'pressure', even after accounting for the potentially confounding shared familial effects (both genetic and environmental). This

suggests that children's genetic influence on weight explains part of the observed phenotypic associations. Our twin analysis provided quantitative estimates of the total variance in parental feeding practices explained by children's genotype. Heritability was substantial for both 'restriction' (43%) and 'pressure' (54%), indicating that children's genes explain about half of the variation in parental feeding behaviour. Multivariate twin analysis established the extent to which parental feeding behaviour was determined by children's genetic influence on BMI specifically. The genetic correlations between children's BMI and both 'restriction' (r_A = 0.28) and pressure (r_A = -0.48) were moderate, indicating overlap between the genes that influence parental feeding behaviour and children's BMI.

A potential confounder of the association between child GPS and parental feeding behaviour, was the parent's own genetic propensity to a higher or lower BMI. Children inherit half of each of their parents' genetic material, so the expected correlation between a child's GPS with that of their parent's is 0.50. A parent's genetic predisposition to be of a higher or lower BMI may also influence the way they feed their children, which could introduce a passive (rather than 'evocative') gene-environment correlation. For example, a parent with a higher BMI may be more restrictive over their child's food intake, but their child also inherits their parent's susceptibility to be of a higher BMI. Restrictive feeding may therefore simply be a marker for a child's genetic predisposition to be of a higher BMI that is transmitted to them by their parent, rather than a causal risk factor (the same could be true for a more pressuring feeding style and lower BMI). In line with this, parental BMI (indexing parental GPS) was significantly positively associated with parental restriction indicating that parents of a higher weight exert greater restriction over their children's food intake (β = 0.08); although the association with parental pressure was not significant. Adjustment for parental BMI did not attenuate the associations between child GPS and either restriction or pressure, suggesting it was not confounding the relationship between parental feeding behaviour and child BMI GPS. Nevertheless, adjustment for parental BMI cannot completely remove confounding from parental BMI, nor can it account for the potential effect of longer-term BMI on parental feeding behaviours. However, in order to rule out confounding by any parental characteristics (both genetic and environmental), we took advantage of a family fixed-effect design, which held the effects of family constant while testing the association between the child BMI GPS and parental feeding practices in DZ co-twins. The within-family analysis allowed us to demonstrate that even after accounting for all genetic and environmental familial effects, parents vary their feeding behaviour for each child depending on their GPS–larger GPS differences between pairs were associated with more pronounced differences in parental feeding behaviour. The magnitudes of the between- and within-family associations between parental feeding behaviour and child GPS were virtually the same, with the exception of the relationship between child GPS and 'restriction' in same-sex twins, strengthening the evidence that children evoke parental responses based on their genetic predispositions for BMI. Nevertheless, as expected, and consistent with the small amount of variance explained in BMI by the GPS, the size of the associations between the BMI GPS and PFPs were small.

Other relevant research

The findings from this study accord with those from twin studies of many other types of parenting behaviours that have also tended to show moderate heritability. A meta-analysis of 32 child twin studies on maternal positivity, negativity, affect and control in relation to parenting showed an average heritability of 24%[18], indicating widespread, child-driven genetic influences on parental behaviour. The heritability estimates for 'restriction' (43%) and 'pressure' (54%) were somewhat higher than the average heritability estimate for the parenting styles

considered in the meta-analysis (24%), but in keeping with the magnitude of the heritability of negative parenting styles observed across early childhood (~55%)[21].

In addition to providing evidence for gene-environment correlation, results from the MZ discordance design also indicated that non-shared environmental influences for child BMI and PFPs are correlated as well. This suggests that child BMI and PFPs are also related due to common non-shared environmental influences. However, the MZ discordance design was not able to shed light on the causal direction–i.e. if child BMI causes PFPs or if PFPs cause child BMI–because our variables were measured at the same time. The few prospective studies that have attempted to establish the cause-effect relationship in the parent-child dynamic using bidirectional analyses have suggested either only a small effect of restriction and/or pressure on child weight, or none[9–11,13]. Prospective studies therefore suggest that PFPs may be less important than is commonly assumed. The well-established strong genetic influence on children's weight–in the order of 70–80%[15,16]–also supports the hypothesis that parents influence child weight via genetic inheritance more than by creating an 'obesogenic' family environment. However, it cannot be ruled out that genetic effects related to BMI in the parents also contribute to an obesogenic environment if gene-environment correlation was at play, further passively reinforcing the child's inherited genetic propensities. The shared environmental influence on BMI in late childhood is also low[15,16]. In the current study, the shared environmental influence on parental feeding behaviour was the proportion of variance that was common to both twins in a pair (invariant within families). It therefore likely reflects variation in feeding behaviour that was parent-driven rather than child-directed. These estimates indicated that a substantial proportion of variation in both 'restriction' (C = 43%) and 'pressure' (C = 37%) also originated in the parent.

Experimental studies in the form of large well-designed randomised controlled trials (RCTs) are needed to truly test the hypothesis that PFPs causally modify children's weight gain trajectories. Very few of these have been conducted to date, and they have focused on the preschool years. Nevertheless, two landmark studies have indicated that parental behaviour may, in fact, be influential in early life. NOURISH[22] was an Australian RCT that randomised 352 parents and infants to receive a feeding intervention (including using low amounts of pressure, and employing child-responsive methods of food restriction) during the period of complementary feeding; 346 families were randomised to the standard care control group. At three to four years of age, children in the intervention group had better appetite control than those in the control group, and there were fewer children with overweight; although this did not reach statistical significance[23]. INSIGHT[24], a US RCT, randomised 145 new mothers to a responsive parenting intervention that focused on feeding infants only in response to their hunger and satiety signals (but neither pressuring nor restricting their milk and food intake), during milk-feeding and complementary feeding; 145 mothers were randomised to a control group. At one year significantly fewer infants in the intervention group had overweight (6%) compared to the control group (13%). These RCTs indicate that parental feeding behaviour can modify young children's eating behaviour and weight gain. However, these studies were conducted in infants and young preschool children so it is unclear whether these findings are generalisable to older children.

The genetic correlations between children's BMI and parental feeding behaviour were modest, and were far from complete (i.e. less than 1.0), indicating that other genetically-determined child characteristics are also influencing parental feeding behaviour. Children's appetite is under strong genetic control; twin studies–including this sample–have shown high heritability for appetite[25,26] and shared heritability with BMI[27]. Appetite is associated with the BMI GPS in this sample and has been shown to mediate part of the GPS-BMI association[28]. It is therefore likely that child appetite also influences parental feeding behaviour[25,26]. In support of this, prospective and within-family studies have provided evidence that within the

context of parental feeding, parents respond not only to their child's weight but also to their eating styles. A large prospective population-based study used bidirectional analyses to show that parents whose children were excessively fussy at baseline increased their pressure over time[29]. A reverse relationship also pertained, but the temporal association from child to parent was stronger. A large within-family study of preschool twins showed that parents varied their pressuring feeding style when their twins were discordant for food fussiness[30]. The fussier twin was pressured more than their co-twin, also in support of a child-driven model of parental feeding behaviour. It stands to reason that a child who is a picky eater is pressured to try some of their vegetables or to eat more overall. Along the same lines, a natural response from a parent who has a child who shows a tendency toward excess intake and a relatively pronounced preference for foods rich in sugar or fat, is to enforce some restriction.

We also found a positive phenotypic correlation between 'restriction' and 'pressure' ($\beta = 0.15$), indicating that parents who exert higher levels of restriction on their children also tend to pressure them more. This suggests that some parents have a more controlling feeding style in general.

Implications and future research

The relationship between parental behaviour and children's emerging characteristics appears to be reciprocal and complex. The current findings suggest that parents' natural feeding responses to child weight are to exert greater restriction of food intake on children with a higher BMI, and to pressure a thinner child to eat. However, these strategies may not be effective in the long run. RCTs have suggested that PFPs can have a lasting and important impact on children's weight and eating behaviour in the early years, although whether or not these findings apply to older children has yet to be determined. It is well established that genetic influence on BMI in younger children is lower, and the shared environmental effect is higher, than it is in older children[15,16]. This suggests that parental influence diminishes as children grow older, gain independence and spend increasing time outside the home with peers rather than parents[31]. Large RCTs that follow children from early life to later childhood are needed to establish if PFPs influence the weight of older children.

Strengths & limitations

A strength of this study is that we used several genetically sensitive methodological approaches to explore the directionality of relationships between child BMI and PFPs, yielding consistent results. PFPs were measured using the Child Feeding Questionnaire, which has well established criterion and construct validity, as well as good internal and test-retest reliability[32]. This instrument has been used widely in previous research into child weight, allowing the findings from this study to be directly compared to a wealth of existing results.

A potential limitation is that heritability estimates from twin studies rely on the assumption that MZs and DZs share their environment in terms of the trait in question to the same extent, so-called the 'equal environments assumption'; if this is violated, the findings are invalid. Therefore if parents feed MZs more similarly than DZs simply because they are identical, this would artificially inflate the MZ correlation and, consequently, heritability. However, if MZs are fed more similarly than DZs because parents are responding to their genetically determined BMI or traits that share genetic influence with BMI such as appetite, differences in feeding experience across MZs and DZs do not constitute a violation of the equal environments assumption because these differences in feeding practices are being driven by greater genetic similarity between MZs than DZs. In addition, if parents' reports of how similarly they fed their twins were biased by their perceived zygosity (i.e. reported treatment was not a true reflection of actual treatment, but related to the twins being MZ or DZ), this would also render

the heritability estimates unreliable. However, this seems unlikely given previous findings that parents' reports about their twins' are not biased by their beliefs about their zygosity, using the 'mistaken zygosity' design[33].

Another limitation was the lack of parental genotypes assessments. Parental BMI is by no means a perfect proxy for their genotypic predisposition to higher or lower BMI; the most powerful approach would be to have parental genotypes whereby the non-transmitted alleles from the parents (which relate to their own BMI and behaviour, but not to that of their child) can be entirely separated from the child's genotype[34]. Nevertheless, the within-family analysis controlled for all family-level genetic and environmental effects, and the magnitudes of the relationships between child BMI and PFPs were unaffected. A further limitation is that we were unable to validate self-reported parental BMI, which may have been inaccurate and could potentially bias our results. Additionally, it may be possible that PFPs are largely explained by environmental factors that influence children's BMI. As the BMI GPS is not yet strong enough to be a sufficient proxy to separate genetic and environmental effects on child BMI, we were unable to test this question empirically. However, considerable genetic correlations between child BMI and PFPs derived from the twin model renders this explanation unlikely. Lastly, BMI was only reported at one time point, but PFPs are likely to be driven by the child's emerging BMI throughout the developmental years. However, BMI-associated SNPs and BMI GPS are associated with weight gain trajectories from infancy throughout childhood, so the BMI GPS in fact captures a long window of child BMI[14,35].

Conclusion

This study provides new evidence for gene-environment correlation in parental feeding practices. We have shown that parental feeding practices are substantially heritable and appear to be partly elicited by the common genetic variants that influence children's BMI. Genome-wide polygenic scores that index children's genetic propensities for their BMI significantly predicted their parents' feeding practices, even after potentially confounding shared family effects were taken into account. The findings of this study provide a new perspective on the nature of the associations between parental feeding practices and child BMI.

Methods

Sample

Participants were drawn from the Twins Early Development Study (TEDS). Between 1994–1996 TEDS recruited over 15,000 twin pairs born in England and Wales, who have been assessed in multiple waves across their development up until the present date. Despite some attrition, about 10,000 twin pairs still actively contribute to TEDS, providing genetic, cognitive, psychological and behavioural data. TEDS participants and their families are representative of families in the UK[36]. Written informed consent was obtained from parents prior to data collection. Project approval was granted by King's College London's ethics committee for the Institute of Psychiatry, Psychology and Neuroscience (05.Q0706/228). This study included 4,445 unrelated individuals with genotyping for the GPS analysis, 2,164 genotyped dizygotic (DZ) twin pairs (1,151 same-sex DZ pairs, 1,013 opposite-sex DZ pairs), and 4,375 twin pairs for the twin analysis (1,636 monozygotic (MZ) pairs, 1,441 same-sex DZ pairs, and 1,298 opposite-sex DZ pairs).

Genotyping

Two different genotyping platforms were used because genotyping was undertaken in two separate waves, five years apart. AffymetrixGeneChip 6.0 SNP arrays were used to genotype 3,665

individuals at Affymetrix, Santa Clara (California, USA) based on buccal cell DNA samples. Genotypes were generated at the Wellcome Trust Sanger Institute (Hinxton, UK) as part of the Wellcome Trust Case Control Consortium 2 (https://www.wtccc.org.uk/ccc2/). Additionally, 8,122 individuals (including 3,607 dizygotic co-twin samples) were genotyped on HumanOmniExpressExome-8v1.2 arrays at the Molecular Genetics Laboratories of the Medical Research Council Social, Genetic Developmental Psychiatry Centre, using DNA that was extracted from saliva samples. After quality control, 635,269 SNPs remained for Affymetrix-GeneChip 6.0 genotypes, and 559,772 SNPs for HumanOmniExpressExome genotypes.

Genotypes from the two platforms were separately phased using EAGLE2[37], and imputed into the Haplotype Reference Consortium (release 1.1) through the Sanger Imputation Service [38] before merging genotype data from both platforms. Genotypes from a total of 10,346 samples (including 3,320 dizygotic twin pairs and 7,026 unrelated individuals) passed quality control, including 3,057 individuals genotyped on Affymetrix and 7,289 individuals genotyped on Illumina. The final data contained 7,363,646 genotyped or well imputed SNPs (for more details, see Supplementary S1 Methods).

We performed principal component analysis on a subset of 39,353 common (MAF > 5%), perfectly imputed (info = 1) autosomal SNPs, after stringent pruning to remove markers in linkage disequilibrium ($r^2 > 0.1$) and excluding high linkage disequilibrium genomic regions so as to ensure that only genome-wide effects were detected.

Phenotypic measures

The samples used for the analyses differed by necessity in order to accommodate the different methodological approaches: unrelated genotyped individuals (UG); dizygotic genotyped co-twins (DG); twin sample (TS) for quantitative genetic analysis. For the classical twin model approach, only phenotypic data from genotyped twins and their co-twins were selected for comparability across the study samples. Descriptive statistics for all phenotypic measures are reported in Supplementary S4A Table for unrelated genotyped individuals, in Supplementary S4B Table for genotyped DZ twin pairs and in Supplementary S4C Table for samples used for twin modelling.

Children's body mass index (BMI) was calculated from parent-reported weight (kg) divided by the square of parent-reported height (metres): kg/m^2. The 1990 UK growth reference data [39] were used to create BMI standard deviation scores (BMI-SDS) which take account of the child's age and sex, and represent the difference between a child's BMI and the mean BMI of the reference children of the same age and sex. BMI-SDS are used rather than BMI itself because BMI varies substantially by age and sex until early adulthood. Reference BMI-SDS have a mean of 0 and a SD of 1: a value greater than 0 indicates a higher BMI than the mean in 1990; a value less than 0 indicates a lower BMI than the mean in 1990. The validity of parent-reported height and weight was tested through home-visits of researchers in a subset of 228 families. Correlations between measurements taken by parents and researchers were high (height: $r = 0.90$; weight: $r = 0.83$)[40]. BMI-SDS were available for 4,259 (UG), 4,134 (DG), and 8,406 (TS) individuals. Children had a mean age of 9.91 years ($SD = 0.87$) when anthropometric measures were assessed.

Parental BMI was calculated for 4,112 individuals using self-reported weight (kg) and height (metres) of the responding parent (kg/m^2), which was assessed at the same time as childhood height and weight. To account for the gender of the responding parent (97% mothers, 3% fathers), we used the z-standardized residuals of gender-corrected BMI in analyses.

To assess PFPs, we used the Child Feeding Questionnaire[41], which parents completed when their twins were approximately 10 years old (mean = 9.91 years, $SD = 0.87$). To measure

the degree to which parents restricted their children's food intake ('restriction'), we calculated a mean composite score based on 6 items (Cronbach's alpha = 0.78), such as "I intentionally keep some foods out of my child's reach", or "If I did not guide my child's eating, he/she would eat too many junk foods". Data were available for 4,386 (UG), 4,228 (DG) and 8,582 (TS) children. Similarly, we created a mean composite score to assess the amount of pressure parents exerted on their children to increase their food intake ('pressure'), including 4 items (Cronbach's alpha = 0.61) such as "If my child says "I'm not hungry", I try to get him/her to eat anyway", or "I have to be especially careful to make sure my child eats enough". Data were available for 4,445 (UG), 4,328 (DG) and 8,750 (TS) children. All items were scored on a 5-point Likert scale (Disagree, Slightly disagree, Neutral, Slightly agree, Agree).

Phenotypic exclusions

For child and parent anthropometrics we removed extreme outliers with implausible values that were deemed to be errors. For children we excluded values based on the following criteria: -/+ 5 standard deviations above or below the mean of height SDS, weight SDS or BMI-SDS; shorter than 105 cm or taller than 180cm; lighter than 12 kg or heavier than 80 kg. After removing outliers, child BMI-SDS had a mean of 0 and a standard deviation of 0.99, showing that the sample is representative of the UK reference population for BMI in 1990 (mean = 0; SD = 1). For parental BMI, we removed individuals with values that fell -/+ 3.5 standard deviations above or below the mean, as well as individuals that weighed below 35 kg. To account for the positive skew, we log transformed this variable. As all variables showed age or sex effects (described in Supplementary S4A, S4B and S4C Table), we controlled for these variables by applying the regression method, using z-standardized residuals for all further analyses. Supplementary S5A, S5B and S5C Table show descriptive statistics for all clean measures (regressed onto age and sex) in unrelated samples, for DZ twin pair samples, and individuals used for twin modelling, respectively.

Genotypic measures

We created Genome-wide Polygenic Scores (GPS) for BMI, using summary statistics from a genome-wide meta-analysis of BMI including 339,224 participants[19]. We calculated a GPS for each individual as the sum of the weighted count of BMI-increasing alleles:

$$GPS_{BMI} = \sum_{i=1}^{k} \beta_i SNP_i$$

where $i \in \{1,2,..,k\}$ and indexes SNP_i and the i number of the k BMI increasing alleles included in the score is determined by the p-value threshold of the SNP–phenotype association in the discovery GWAS, the β-coefficients for each respective genetic variant is used as a weight, and the count of each reference allele is represented by genotype dosage (0,1, or 2 alleles) of SNP_i.

We used the software PRSice[42] to calculate GPS in our sample. To account for multicollinearity among SNPs in Linkage Disequilibrium (LD), which can upwardly bias GPS predictions[43], genome-wide clumping was performed ($r^2 = 0.1$, kb = 250). Using the clumped, independent SNPs, we created eight GPS for 10,346 individuals (7,026 unrelated individuals; 3,320 DZ twin pairs) using increasingly liberal GWAS p-value thresholds (pT: 0.001,0.05, 0.1,0.2,0.3,0.4,0.5,1). Diagonals in Supplementary S2 Fig show the number of SNPs included in each respective GPS. As all thresholds performed similarly well (Supplementary S2 Fig), we used a GPS based on the smallest p-value threshold of 0.001 for all further analyses. Potential effects due to population stratification and genotyping were accounted for by regressing the first ten principal components, and factors capturing genotyping information (microarray,

batch and plate) onto the child BMI GPS, subsequently using the z-standardised residuals in our analyses.

Statistical analysis: Genome-wide polygenic score (GPS) analyses

Trait prediction in unrelated samples. Associations between child BMI GPS and phenotypes were assessed using linear regression analyses. All variables were standardised prior to analyses, therefore β coefficients from linear regression models are equivalent to Pearson's correlation coefficients.

Within-family analyses: Accounting for family effects in unrelated samples and DZ twin pairs. Children not only inherit half of each of their parent's DNA, but also the family environment. Therefore, it is possible that familial effects, both genetic and environmental, confound the relationships between child GPS and PFPs. To account for these potential confounding effects, we used two approaches. Firstly, we removed variance in the PFPs (restriction, pressure) explained by parental BMI in our sample of unrelated individuals using the regression method, and repeated association analyses. Secondly, we used data on genotyped DZ twin pairs to explicitly model the effect of within-DZ twin pair GPS differences on differences in PFPs by accounting for the family contributions in a fixed-effects model:

$$Y_{ij} = \alpha_j + \beta GPS_{ij} + e_{ij},$$

where i ∈ {1,2} indexes the individuals of the dizygotic twin pairs, and j ∈ {1,2,..,k} indexes the k families (i.e. sets of dizygotic twin pairs). Thus, Y_{ij} is the trait value for the *i*th individual of the *j*th family, α_j is a vector including the (fixed) family effects, β is the effect of the GPS within families, e_{ij} is the random error for each individual and each family with $e_{ij} \sim N(0,\sigma^2)$, and Cov ($\alpha_j$, e_{ij}) = 0. The family units were coded using dummy variables in order to estimate the α_j effects. By accounting for the differences in contributing factors between families via α_j, this model tests for the effect of differences in GPS values between DZ twins on the outcome and therefore assesses the impact of GPS with shared genetic and shared environmental factors accounted for. The within-family associations indicate the extent to which parents vary their 'restriction' or 'pressure' in response to differences in their twins' BMI GPS. A larger association indicates that the greater the difference between twin pairs' BMI GPS, the greater the difference in parental 'restriction' or 'pressure' across two twins in a pair. We applied fixed-effects models to our combined DZ data, and repeated these analyses using same-sex DZ pairs and opposite-sex DZ samples only.

Statistical analysis: Twin modelling

To obtain broad estimates of the extent to which individual differences in PFPs are determined by children's genotypes, we used a multivariate 'correlated factors' twin model. This allowed us to estimate: (1) the heritability of PFPs, which provided an indication of the extent to which PFPs are caused by children's genotypes in general; and (2) the extent of common genetic influence on both child BMI-SDS and PFPs, which provided an indication of the extent to which PFPs are caused by children's genetic propensity to higher or lower BMI, specifically.

Based on biometrical genetics theory[44], it is possible to decompose variance in a single trait into three components: additive genetic (A; heritability), shared environmental (C; all environmental effects that make family members more similar) and non-shared environmental (E; all environmental effects that contribute to dissimilarities across family members, including random error measurement). The basis of the method is to compare resemblance for a single trait between monozygotic (MZ) and dizygotic (DZ) twin pairs, who share 100% and 50% (on average) of their segregating genetic material, respectively, while both types of

twins are correlated 100% for their shared environmental influence. The observed covariation between MZ and DZ pairs is compared with the expected covariation, based on the knowledge of different degrees of allele sharing (or identity by descent (IBD)) of MZ (IBD = 1.0) and DZ pairs (IBD = 0.5 on average). The twin method therefore assumes that MZ and DZ twins share their environments in terms of the trait in question to the same extent (so-called the 'equal environments assumption'), and the only difference between the two types of twins is the extent of their genetic relatedness.

Using the same principles, comparison of MZ and DZ covariation across traits (so-called cross-twin cross-trait covariance, e.g. the covariation between twin 1 BMI-SDS and twin 2 'restriction') provides an indication of the extent to which the genetic and environmental influences on multiple traits are the same. The key pieces of information provided are the aetiological correlations, which indicate the extent to which child BMI and PFPs are caused by the same additive genetic (genetic correlation; r_A), shared environmental (shared environmental correlation; r_C), and non-shared environmental influences (non-shared environmental correlation; r_E). In this analysis we were primarily interested in the genetic correlation, which indicates the extent to which the additive genetic influences on child BMI cause PFPs. The aetiological correlations range from -1 to 1 and can be interpreted similarly to Pearson's correlations. For example, a high positive genetic correlation between 'restriction' and BMI would indicate that many of the DNA variants that cause higher child BMI are the same as those cause higher levels of 'restriction', while a high negative genetic correlation would indicate that many of the DNA variants causing higher child BMI are the same as those causing lower levels of 'restriction'.

Maximum likelihood structural equation modelling was used to estimate intra-class correlations across the zygosities, the A, C and E parameter estimates and aetiological correlations (with 95% confidence intervals), and goodness-of-fit statistics. Sex differences in the parameter estimates were also tested for using a sex-limitation model. Analyses were implemented in the R package *OpenMx*[45].

Monozygotic twin discordance analysis. MZ twins share 100% of their genotypic information and grow up in the same family, suggesting that phenotypic differences that are not due to measurement error are caused by non-shared environmental influences; because they cannot be explained by genetic or shared environment differences[46]. In order to identify non-shared environmental sources of PFP in relation to child BMI, we calculated within MZ pair difference scores for child BMI-SDS, 'restriction' and 'pressure' for all MZ pairs by subtracting the variable score for twin 2 from the variable score of twin 1 (for variable distributions and descriptive statistics, see Supplementary S2 Fig). Therefore, the twin difference score is evaluated in respect to twin 1 (e.g. a positive value indicates that twin 1 has a higher value than twin 2). We applied linear regression analysis to identify whether within MZ twin pair differences in BMI-SDS predicted MZ twin differences in PFPs.

Supporting information

S1 Methods. Genotyping and quality control.
(DOCX)

S1 Table. Fit statistics for the multivariate model including child BMI SDS, parental pressure and parental restriction. ep = estimated parameters; -2LL = -2 log likelihood; df = degrees of freedom; AIC = Akaike Information Criterion. An ACE model without scalar (no sex-limitation) provided best fit.
(XLSX)

S2 Table. Twin intra-class correlations by sex and zygosity groups. ACE estimates are based on no sex-limitation models. 95% confidence intervals are shown in square brackets.
(XLSX)

S3 Table. Phenotypic, genetic, shared environmental, non-shared environmental correlations and 95% confidence intervals. r_P = Phenotypic correlation; r_A = genetic correlation; r_C = shared environmental correlation; r_E = non-shared environmental correlation. All estimates are based on maximum likelihood.
(XLSX)

S4 Table. Raw descriptive statistics of phenotypic measures in (a) genotyped unrelated individuals, (b) genotyped DZ twins, (c) twins for twin modelling by zygosity.
S4A Table: Discrepancies in sample sizes between height and weight and their respective SD scores is due to list wise deletion in the construction of SD scores due to missing age. *F* = *F*-statistic of ANOVA. R^2 = Variance explained. S4B Table: The sample includes all DZ pairs (same sex and opposite sex); the N includes the total number of individual DZs. Discrepancies in sample sizes between height and weight and their respective SD scores is due to list wise deletion in the construction of SD scores due to missing age. Twin pairs with incomplete data were excluded. *F* = *F*-statistic of ANOVA (performed on one randomly selected twin per pair). R^2 = Variance explained. S4C Table: Means for phenotypic measures and standard deviations in brackets. MZ = monozygotic; DZ = dizygotic; m = male; f = female; os = opposite sex. Twin pairs with incomplete data and missing information about zygosity were excluded, which explains slight sample size deviations in comparison to unrelated genotyped samples. *F*-statistics reported for sex, zygosity and sex*zygosity interaction. R^2 = variance explained by sex, zygosity and their interaction (ANOVA). All Tables: * = $p<0.05$; ** = $p<0.01$; *** = $p<0.001$.
(XLSX)

S5 Table. Descriptive statistics of cleaned phenotypic measures (regressed onto age and sex) in (a) genotyped unrelated individuals, (b) genotyped DZ twins, (c) twins for twin modelling by zygosity.
S5A Table: Discrepancies in sample sizes between height and weight and their respective SD scores is due to list wise deletion in the construction of SD scores due to missing age. *F* = *F*-statistic of ANOVA. R^2 = Variance explained. S5B Table: The sample includes all DZ pairs (same sex and opposite sex); the N includes the total number of individual DZs. Discrepancies in sample sizes between height and weight and their respective SD scores is due to list wise deletion in the construction of SD scores due to missing age. Twin pairs with incomplete data were excluded. *F* = *F*-statistic of ANOVA (performed on one randomly selected twin per pair). R^2 = Variance explained. S5C Table: Means for phenotypic measures and standard deviations in brackets. MZ = monozygotic; DZ = dizygotic; m = male; f = female; os = opposite sex. Twin pairs with incomplete data and missing information about zygosity were excluded, which explains slight sample size deviations in comparison to unrelated genotyped samples. *F*-statistics reported for sex, zygosity and sex*zygosity interaction. R^2 = variance explained by sex, zygosity and their interaction (ANOVA). All Tables: * = $p<0.05$; ** = $p<0.01$; *** = $p<0.001$.
(XLSX)

S1 Fig. MZ twin difference score distributions and descriptive statistics.
(TIF)

S2 Fig. Correlations across all GPS and phenotypic measures. Diagonals of Genome-wide Polygenic Scores (GPS) show number of SNPs included in each respective score. * = $p<0.05$;

** = $p<0.01$; *** = $p<0.001$.
(TIF)

Author Contributions

Conceptualization: Saskia Selzam, Robert Plomin, Clare H. Llewellyn.

Data curation: Saskia Selzam, Jonathan R. I. Coleman.

Formal analysis: Saskia Selzam.

Funding acquisition: Robert Plomin.

Investigation: Saskia Selzam, Robert Plomin, Clare H. Llewellyn.

Methodology: Saskia Selzam, Tom A. McAdams, Paul F. O'Reilly.

Project administration: Clare H. Llewellyn.

Supervision: Robert Plomin, Clare H. Llewellyn.

Visualization: Saskia Selzam.

Writing – original draft: Saskia Selzam, Clare H. Llewellyn.

Writing – review & editing: Saskia Selzam, Tom A. McAdams, Jonathan R. I. Coleman, Susan Carnell, Paul F. O'Reilly, Robert Plomin, Clare H. Llewellyn.

References

1. Clark HR, Goyder E, Bissell P, Blank L, Peters J. How do parents' child-feeding behaviours influence child weight? Implications for childhood obesity policy. J Public Health. 2007 Jun 1; 29(2):132–41.
2. Golan M. Parents as agents of change in childhood obesity - from research to practice. Pediatric Obesity. 2006 Jan 1; 1(2):66–76.
3. Lindsay AC, Sussner KM, Kim J, Gortmaker S. The role of parents in preventing childhood obesity. Future Child. 2006 Apr 1; 16(1):169–86. PMID: 16532663
4. Birch LL, Fisher JO. Development of Eating Behaviors Among Children and Adolescents. Pediatrics. American Academy of Pediatrics; 1998 Mar 1; 101(Supplement 2):539–49.
5. Birch LL, Fisher JO. Mothers' child-feeding practices influence daughters' eating and weight. Am J Clin Nutr. 2000 May; 71(5):1054–61. https://doi.org/10.1093/ajcn/71.5.1054 PMID: 10799366
6. Batsell WR, Brown AS, Ansfield ME, Paschall GY. "You will eat all of that!": A retrospective analysis of forced consumption episodes. Appetite. 2002 Jun; 38(3):211–9. PMID: 12071687
7. Galloway AT, Fiorito LM, Francis LA, Birch LL. "Finish your soup": Counterproductive effects of pressuring children to eat on intake and affect. Appetite. 2006 May 1; 46(3):318–23. https://doi.org/10.1016/j.appet.2006.01.019 PMID: 16626838
8. Shloim N, Edelson LR, Martin N, Hetherington MM. Parenting Styles, Feeding Styles, Feeding Practices, and Weight Status in 4–12 Year-Old Children: A Systematic Review of the Literature. Front Psychol. 2015 Dec; 6:1849. https://doi.org/10.3389/fpsyg.2015.01849 PMID: 26696920
9. Afonso L, Lopes C, Severo M, Santos S, Real H, Durão C, et al. Bidirectional association between parental child-feeding practices and body mass index at 4 and 7 y of age. Am J Clin Nutr. 2016 Feb 3; 103(3):861–7. https://doi.org/10.3945/ajcn.115.120824 PMID: 26843159
10. Derks IP, Tiemeier H, Sijbrands EJ, Nicholson JM, Voortman T, Verhulst FC, et al. Testing the direction of effects between child body composition and restrictive feeding practices: results from a population-based cohort. Am J Clin Nutr. 2017 Aug 9; 106(3):783–90. https://doi.org/10.3945/ajcn.117.156448 PMID: 28793987
11. Jansen PW, Tharner A, van der Ende J, Wake M, Raat H, Hofman A, et al. Feeding practices and child weight: is the association bidirectional in preschool children? Am J Clin Nutr. 2014 Sep 3; 100(5):1329–36. https://doi.org/10.3945/ajcn.114.088922 PMID: 25332330
12. Rhee KE, Coleman SM, Appugliese DP, Kaciroti NA, Corwyn RF, Davidson NS, et al. Maternal Feeding Practices Become More Controlling After and Not Before Excessive Rates of Weight Gain. Obesity. 2009 Sep 1; 17(9):1724–9. https://doi.org/10.1038/oby.2009.54 PMID: 19282827

13. Webber L, Cooke L, Hill C, Wardle J. Child adiposity and maternal feeding practices a longitudinal analysis. Am J Clin Nutr. 2010 Sep 29; 92(6):1423–8. https://doi.org/10.3945/ajcn.2010.30112 PMID: 20881070

14. Elks CE, Hoed den M, Zhao JH, Sharp SJ, Wareham NJ, Loos RJF, et al. Variability in the heritability of body mass index: a systematic review and meta-regression. Front Endocrinol. 2012 Feb 28; 3:29.

15. Silventoinen K, Jelenkovic A, Sund R, Hur Y-M, Yokoyama Y, Honda C, et al. Genetic and environmental effects on body mass index from infancy to the onset of adulthood: an individual-based pooled analysis of 45 twin cohorts participating in the COllaborative project of Development of Anthropometrical measures in Twins (CODATwins) study. Am J Clin Nutr. 2016 Jul 13; 104(2):371–9. https://doi.org/10.3945/ajcn.116.130252 PMID: 27413137

16. Silventoinen K, Rokholm B, Kaprio J, Sørensen TIA. The genetic and environmental influences on childhood obesity: a systematic review of twin and adoption studies. International Journal of Obesity. 2010 Jan 1; 34(1):29–40. https://doi.org/10.1038/ijo.2009.177 PMID: 19752881

17. McGuire S. The Heritability of Parenting. Parenting: Science And Practice. 2003 Feb 1; 3(1):73–94.

18. Avinun R, Knafo A. Parenting as a Reaction Evoked by Children's Genotype: A Meta-Analysis of Children-as-Twins Studies. Personality and Social Psychology Review. 2014 Feb; 18(1):87–102. https://doi.org/10.1177/1088868313498308 PMID: 23940232

19. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, et al. Genetic studies of body mass index yield new insights for obesity biology. 2015 Feb; 518(7538):197–206.

20. Krapohl E, Euesden J, Zabaneh D, Pingault J-B, Rimfeld K, Stumm von S, et al. Phenome-wide analysis of genome-wide polygenic scores. Molecular Psychiatry 2016 Sep; 21(9):1188–93. https://doi.org/10.1038/mp.2015.126 PMID: 26303664

21. Knafo A, Plomin R. Parental discipline and affection and children's prosocial behavior: genetic and environmental links. J Pers Soc Psychol. 2006 Jan; 90(1):147–64. https://doi.org/10.1037/0022-3514.90.1.147 PMID: 16448315

22. Daniels LA, Magarey A, Battistutta D, Nicholson JM, Farrell A, Davidson G, et al. The NOURISH randomised control trial: Positive feeding practices and food preferences in early childhood—a primary prevention program for childhood obesity. BMC Public Health. 2009 Dec 1; 9(1):387.

23. Magarey A, Mauch C, Mallan K, Perry R, Elovaris R, Meedeniya J, et al. Child Dietary and Eating Behavior Outcomes up to 3.5 Years After an Early Feeding Intervention: The NOURISH RCT. Obesity. 2016 Jul; 24(7):1537–45. https://doi.org/10.1002/oby.21498 PMID: 27193736

24. Paul IM, Williams JS, Anzman-Frasca S, Beiler JS, Makova KD, Marini ME, et al. The Intervention Nurses Start Infants Growing on Healthy Trajectories (INSIGHT) study. BMC Pediatr. 2014 Dec; 14(1):184.

25. Llewellyn CH, van Jaarsveld CHM, Johnson L, Carnell S, Wardle J. Nature and nurture in infant appetite: analysis of the Gemini twin birth cohort. Am J Clin Nutr. 2010 Mar 24; 91(5):1172–9. https://doi.org/10.3945/ajcn.2009.28868 PMID: 20335548

26. Carnell S, Haworth CMA, Plomin R, Wardle J. Genetic influence on appetite in children. International Journal of Obesity. 2008 Oct 1; 32(10):1468–73. https://doi.org/10.1038/ijo.2008.127 PMID: 18679413

27. Llewellyn CH, van Jaarsveld CH, Plomin R, Fisher A, Wardle J. Inherited behavioral susceptibility to adiposity in infancy: a multivariate genetic analysis of appetite and weight in the Gemini birth cohort. Am J Clin Nutr. 2012 Jan 25; 95(3):633–9. https://doi.org/10.3945/ajcn.111.023671 PMID: 22277555

28. Llewellyn CH, Trzaskowski M, van Jaarsveld CHM, Plomin R, Wardle J. Satiety Mechanisms in Genetic Risk of Obesity. JAMA Pediatr. American Medical Association; 2014 Apr 1; 168(4):338–44. https://doi.org/10.1001/jamapediatrics.2013.4944 PMID: 24535189

29. Jansen PW, de Barse LM, Jaddoe VWV, Verhulst FC, Franco OH, Tiemeier H. Bi-directional associations between child fussy eating and parents' pressure to eat: Who influences whom? Physiol Behav. 2017 Jul 1; 176:101–6. https://doi.org/10.1016/j.physbeh.2017.02.015 PMID: 28215424

30. Harris HA, Fildes A, Mallan KM, Llewellyn CH. Maternal feeding practices and fussy eating in toddlerhood: a discordant twin analysis. International Journal of Behavioral Nutrition and Physical Activity. 2016 Dec 1; 13(1):81.

31. Savage JS, Birch LL, Marini M, Anzman-Frasca S, Paul IM. Effect of the INSIGHT Responsive Parenting Intervention on Rapid Infant Weight Gain and Overweight Status at Age 1 Year A Randomized Clinical Trial. JAMA Pediatr. American Medical Association; 2016 Aug 1; 170(8):742–9. https://doi.org/10.1001/jamapediatrics.2016.0445 PMID: 27271455

32. Vaughn AE, Tabak RG, Bryant MJ, Ward DS. Measuring parent food practices: a systematic review of existing measures and examination of instruments. International Journal of Behavioral Nutrition and Physical Activity 2016 13:1. BioMed Central; 2013 Dec; 10(1):61.

33. Herle M, Fildes A, van Jaarsveld C, Rijsdijk F, Llewellyn CH. Parental Reports of Infant and Child Eating Behaviors are not Affected by Their Beliefs About Their Twins' Zygosity. Behav Genet. 2016 Nov 1; 46 (6):763–71. https://doi.org/10.1007/s10519-016-9798-y PMID: 27406596

34. Kong A, Thorleifsson G, Frigge ML, Vilhjalmsson BJ, Young AI, Thorgeirsson TE, et al. The nature of nurture: Effects of parental genotypes. Science. 2018 Jan; 359(6374):424–8. https://doi.org/10.1126/science.aan6877 PMID: 29371463

35. Hardy R, Wills AK, Wong A, Elks CE, Wareham NJ, Loos RJF, et al. Life course variations in the associations between FTO and MC4R gene variants and body size. Hum Mol Genet. 2009 Oct 31; 19(3):545–52. https://doi.org/10.1093/hmg/ddp504 PMID: 19880856

36. Haworth CMA, Davis OSP, Plomin R. Twins Early Development Study (TEDS): A Genetically Sensitive Investigation of Cognitive and Behavioral Development From Childhood to Young Adulthood. Twin Res Hum Genet. 2013 Feb; 16(1):117–25. https://doi.org/10.1017/thg.2012.91 PMID: 23110994

37. Loh P-R, Danecek P, Palamara PF, Fuchsberger C, A Reshef Y, K Finucane H, et al. Reference-based phasing using the Haplotype Reference Consortium panel. Nat Genet. 2016 Nov; 48(11):1443–8. https://doi.org/10.1038/ng.3679 PMID: 27694958

38. McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, et al. A reference panel of 64,976 haplotypes for genotype imputation. Nat Genet. 2016 Oct; 48(10):1279–83. https://doi.org/10.1038/ng.3643 PMID: 27548312

39. Cole TJ, Freeman JV, Preece MA. Body mass index reference curves for the UK, 1990. Archives of Disease in Childhood. 1995 Jul 1; 73(1):25–9. PMID: 7639544

40. Wardle J, Carnell S, Haworth CMA, Plomin R. Evidence for a strong genetic influence on childhood adiposity despite the force of the obesogenic environment. Am J Clin Nutr. 2008 Feb 1; 87(2):398–404. PMID: 18258631

41. Birch LL, Fisher JO, Grimm-Thomas K, Markey CN, Sawyer R, Johnson SL. Confirmatory factor analysis of the Child Feeding Questionnaire: a measure of parental attitudes, beliefs and practices about child feeding and obesity proneness. Appetite. 2001 Jun 1; 36(3):201–10. PMID: 11358344

42. Euesden J, Lewis CM, O'Reilly PF. PRSice: Polygenic Risk Score software. Bioinformatics. 2014 Dec 29; 31(9):1466–8. https://doi.org/10.1093/bioinformatics/btu848 PMID: 25550326

43. Palla L, Dudbridge F. A Fast Method that Uses Polygenic Scores to Estimate the Variance Explained by Genome-wide Marker Panels and the Proportion of Variants Affecting a Trait. Am J Hum Genet. 2015 Aug 6; 97(2):250–9. https://doi.org/10.1016/j.ajhg.2015.06.005 PMID: 26189816

44. Mather K, Jinks JL. Introduction to biometrical genetics. London: Chapman and Hall; 1977.

45. Boker S, Neale M, Maes H, Wilde M, Spiegel M, Brick T, et al. OpenMx: An Open Source Extended Structural Equation Modeling Framework. Psychometrika. 2011 Apr 1; 76(2):306–17. https://doi.org/10.1007/s11336-010-9200-6 PMID: 23258944

46. Pike A, Reiss D, Hetherington EM, Plomin R. Using MZ Differences in the Search for Nonshared Environmental Effects. J Child Psychol Psychiatry. 1996 Sep; 37(6):695–704. PMID: 8894950

Chapter 6 – Comparing within- and between-family polygenic score prediction

This chapter has been adapted from a manuscript that has been accepted for publication in the *American Journal of Human Genetics*:

Selzam, S.[1], Ritchie, S. J. [1], Pingault, J. B. [1,2], Reynolds, C. A.[3], O'Reilly, P. F.[1,4], Plomin, R.[1] (*Accepted*). Comparing within- and between-family polygenic score prediction. *American Journal of Human Genetics.*

[1]Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, SE5 8AF, United Kingdom
[2]Division of Psychology and Language Sciences, University College London, London, WC1H 0AP, United Kingdom
[3]Department of Psychology, University of California Riverside, Riverside, CA 92521, United States of America
[4]Icahn School of Medicine, Mount Sinai, New York City, NY 10029, United States of America

Supplemental materials are included under Appendix 5.

Abstract

Polygenic scores are a popular tool for prediction of complex traits. However, prediction estimates in samples of unrelated participants can include effects of population stratification, assortative mating and environmentally mediated parental genetic effects, a form of genotype-environment correlation (rGE). Comparing genome-wide polygenic score (GPS) predictions in unrelated individuals with predictions between siblings in a within-family design is a powerful approach to identify these different sources of prediction. Here, we compared within- to between-family GPS predictions of eight outcomes (anthropometric, cognitive, personality and health) for eight corresponding GPSs. The outcomes were assessed in up to 2,366 dizygotic (DZ) twin pairs from the Twins Early Development Study from age 12 to age 21. To account for family clustering, we used mixed-effects modelling, simultaneously estimating within- and between-family effects for target- and cross-trait GPS prediction of the outcomes. There were three main findings: (1) DZ twin GPS differences predicted DZ differences in height, BMI, intelligence, educational achievement and ADHD symptoms; (2) target and cross-trait analyses indicated that GPS prediction estimates for cognitive traits (intelligence and educational achievement) were on average 60% greater between families than within families, but this was not the case for non-cognitive traits; and (3) much of this within- and between-family difference for cognitive traits disappeared after controlling for family socio-economic status (SES), suggesting that SES is a major source of between-family prediction through rGE mechanisms. These results provide insights into the patterns by which rGE contributes to GPS prediction, while ruling out confounding due to population stratification and assortative mating.

Introduction

The recent influx of well-powered genome-wide association (GWA) studies has led to substantial advances in our ability to detect genetic associations between single base pair variants (single nucleotide polymorphisms; SNPs) across the genome and a myriad of complex traits. Although individual SNP effect sizes are extremely small (Gratten, Wray, Keller, & Visscher, 2014), the surge in GWA power has improved the ability to predict complex traits through the genome-wide polygenic score (GPS) approach (Martin, Daly, Robinson, Hyman, & Neale, 2018; Plomin & Stumm, 2018). GPSs are indices of individuals' genetic propensity for a trait, and are derived as the sum of the total number of trait-associated alleles across the genome, weighted by their respective association effect size estimated through GWA analysis(Wray et al., 2014). GPS can be calculated in any sample with genotype data that is independent from the discovery GWA study, and have permeated research in the social, behavioural and biomedical sciences (Plomin, 2018). In this paper, we use within-family analysis to investigate an important potential source of prediction in polygenic score analysis: passive genotype-environment correlation.

Currently one of the largest GWA meta-analyses with a sample size of 1.1 million was performed on educational attainment (years of schooling) (Lee et al., 2018). A GPS derived from this study is the most predictive GPS for any behavioural trait to date, explaining 10.6% of the variance in years of education (Lee et al., 2018) and 14.8% in tested educational achievement (Allegrini et al., 2019). The predictive power of the educational attainment GPS (EA GPS) is considerable in contrast to other GPS for behavioural traits. Notably, cross-trait analyses have revealed that EA GPS is widely associated with traits other than educational achievement, including intelligence (Allegrini et al., 2019; Lee et al., 2018; Plomin & Stumm, 2018), socioeconomic status (SES) (Belsky et al., 2018; 2016; Hill et al., 2016; Selzam et al., 2017), behaviour problems (de Zeeuw et al., 2014) , mental health (Hagenaars et al., 2016), physical health (Hagenaars et al., 2016) and personality (Mõttus, Realo, Vainik, Allik, & Esko, 2017; Smith-Woolley, Selzam, & Plomin, 2019), in some cases accounting for as much as or more than the variance in cross-trait associations explained by the target GPS themselves (Krapohl et al., 2016; Smith-Woolley et al., 2019).

However, GWA analyses, and the GPSs derived from them in independent samples, are naïve to the pathways that lead from SNPs to trait outcomes (Belsky & Harden, 2019). With a

focus on prediction, the mechanisms by which polygenic scores relate to phenotypes are left largely unexplored. Given the popularity and widespread use of the GPS approach, the interpretation of GPS prediction estimates requires more careful consideration. Potentially, *passive genotype-environment correlation* (prGE) (Plomin, DeFries, & Loehlin, 1977) effects could be one source of prediction. Parents generate family environments consistent with their own genotypes, which in turn facilitate the development of the offspring trait, thus inducing a correlation between offspring genotype and family environment (Bates et al., 2018; Kong et al., 2018; Pingault et al., 2018). Although these effects are also genetic in origin, they stem from the parents and are thus environmentally mediated. Therefore, GPS prediction among unrelated individuals may include contributions from both direct genetic effects and also indirect effects due to prGE.

Within-family analysis of siblings is a powerful approach to disentangle these potential sources of prediction. The additive genetic correlation between siblings is on average 0.50 (Fisher, 1918), and the transmission of alleles from parents to offspring is randomized during meiosis, such that siblings have equal probability of inheriting any given allele (Fletcher, 2011). The variability around the average genetic relationship between siblings due to random segregation is generally independent of the environment, therefore any genetic difference between siblings is free of shared environmental influence (Young et al., 2018). A relationship between their genetic differences and trait differences provides evidence for a causal effect of the measured genetic difference, since (i) siblings are well-matched on all shared familial genetic influences that shape the environment, and (ii) potential bias due to population stratification and assortative mating is completely eliminated within families (Benyamin, Visscher, & McRae, 2009; Brumpton et al., 2019; Lee et al., 2018). Such within-family analyses account for prGE effects that are related to common family environments which are correlated with the transmitted alleles shared between siblings, but also environmental effects related to non-transmitted parental alleles that contribute to offspring similarity within a family. The use of DZ co-twins strengthens this design further as all shared environmental influences are time-invariant between twins (e.g. pregnancy risk factors, parental age, family income).

Indeed, previous within-family analyses have revealed substantial reductions in individual SNP effect sizes. For example, there was an effect size attenuation of ~40% compared to between-family associations in the most recent GWA study on educational attainment (Lee

et al., 2018). Most of this reduction has been attributed to prGE; no similar deflation of effect sizes was found for height (Lee et al., 2018), indicating that prGE is not likely at play. A novel method relying on close- and distantly-related individuals, and that is applied to very large populations, detected a similar reduction of SNP-heritability estimates of educational achievement (~40%) (Young et al., 2018). Moreover, studies that tested the effect between non-transmitted alleles from parental to offspring genotypes on offspring outcomes reported a significant association for educational attainment (Bates et al., 2018; Kong et al., 2018) – an effect of so-called *genetic nurture* – but not for height and BMI (Bates et al., 2018; Kong et al., 2018). In contrast, one study that tested within-family predictions of educational attainment using the EA GPS found no noteworthy difference in comparison to between-family estimates (Domingue, Belsky, Conley, Harris, & Boardman, 2015). However, this GPS was based on the first GWA study for educational attainment (Rietveld et al., 2014), and may have been underpowered to pick up prGE-driven effects. Indeed, a more recent study found that using the latest GPS for educational attainment, there was an attenuation of ~55% in the prediction of years of schooling within families in comparison to between-family estimates (Trejo & Domingue, 2019).

Overall, relatively little research has been conducted on within-family GPS prediction, mostly focussing on educational and anthropometric traits. This study adds substantially to this literature by systematically comparing within-family GPS prediction to between-family GPS prediction across eight life outcomes (height, BMI, self-rated health, intelligence, educational achievement, neuroticism, attention-deficit/hyperactivity symptoms, and schizophrenia symptoms). Educational achievement is both phenotypically and genetically correlated with many life outcomes (Briley, Domiteaux, & Tucker-Drob, 2014; De Ridder et al., 2013; Krapohl et al., 2014; Marques, Pais-Ribeiro, & Lopez, 2011; Pingault et al., 2011; B. Spinath, Spinath, Harlaar, & Plomin, 2006; Zuffianò et al., 2013). It is also is highly genetically correlated with family SES (Hill et al., 2016; Krapohl & Plomin, 2016; Trzaskowski et al., 2014), and EA GPS predicts 7.3% of the variance in SES (Selzam et al., 2017). Therefore, it is possible that the effects identified in the GWA studies for educational attainment related to family environment (e.g. SES) also contribute to the development of other behavioural traits through prGE mechanisms. Although it has been suggested that the widespread cross-trait associations between the EA GPS and various outcomes may be partly driven by prGE effects (Koellinger & Harden, 2018; Smith-Woolley et al., 2019), to our knowledge no study to date has tested this hypothesis.

It is the aim of this study to investigate potential influences of prGE in a range of life outcomes through the comparison of within- and between-family polygenic score prediction estimates. First, we predict that within-family estimates will be disproportionally lower than between-family estimates for EA GPS predictions of educational achievement in contrast to other GPS predictions of their target trait. Second, we predict that cross-trait associations between the EA GPS and other outcomes will be smaller within families than between families, in comparison to the cross-trait associations of other GPSs.

Methods

Our hypotheses, measures and analysis plan were preregistered with the Open Science Framework (for more details, see Online Resource section), except where indicated below. The non-preregistered analyses should be considered exploratory.

Sample

Participants were drawn from the Twins Early Development Study (TEDS). Between 1994-1996 TEDS recruited 16,810 twin pairs born in England and Wales, who have been assessed in multiple waves across development until the present. The demographic characteristics of TEDS participants and their families closely match those of families in the UK (Haworth, Davis, & Plomin, 2013; Selzam et al., 2017). Written informed consent was obtained from parents prior to data collection, and from TEDS participants themselves past the age of 18. Project approval was granted by King's College London's ethics committee for the Institute of Psychiatry, Psychology and Neuroscience PNM/09/10–104. Only DZ co-twins with complete data were included in this study.

Phenotypic data

Height. Self-reported height was assessed at the average age of 22.1 (SD=0.86) in 1,463 twin pairs.

Body Mass Index (BMI). BMI was calculated using self-reported weight in kg and height in meters ($\frac{kg}{m^2}$) at age 22.1 (SD=0.86) in 1,353 twin pairs.

Self-rated health. Twins rated their health on the reduced RAND Short-Form Health Survey (Ware & Sherbourne, 1992). Individuals scored their health on a five point Likert scale for five questions such as "In general, would you say your health is?" ("Poor" to "Excellent"), or "I am as healthy as anybody I know" ("Strongly Disagree" to "Strongly Agree"). Data were available on 1,494 twin pairs at age 22.1 (SD=0.86).

Intelligence. At age 11.4 (SD=0.65), twins were assessed on their non-verbal abilities (Raven's Standard Progressive Matrices (Raven & Court, 1996); WISC-III-UK Picture Completion (Wechsler, 1992)) and on their verbal abilities (WISC-III-PI Vocabulary Multiple-Choice (Kaplan, Fein, Kramer, Delis, & Morris, 1999); WISC-III-PI Information Multiple-Choice (Kaplan et al., 1999)). A composite variable was calculated as the arithmetic mean of the z-standardized scales for 1,569 twin pairs.

Educational achievement. Results for standardized tests taken at the end of compulsory education in the United Kingdom (General Certificate of Secondary Education; GCSE) were obtained for twins at age 16.3 (SD=0.29) via self-report. Grades were coded from 4 (G; the minimum pass grade) to 11 (A*; the highest possible grade). Self-reported GCSE grades in TEDS highly correlate with grades obtained for a subsample of individuals from the National Pupil Database ($r = 0.98$ for English, $r = 0.99$ for mathematics, $r > 0.95$ for all sciences) (Krapohl et al., 2014). A composite was calculated as the arithmetic mean of the compulsory core subjects – Maths, English and Science – for 2,366 twin pairs.

Neuroticism. At age 16.5 (SD=0.27), twins were assessed on their Big Five personality traits on a five-point Likert scale (Mullins-Sweatt, Jamerson, Samuel, Olson, & Widiger, 2006). For this study, we used the six Neuroticism items (e.g. Anxiousness; Vulnerability) to form a composite score by taking the arithmetic mean for 789 twin pairs.

Attention-Deficit Hyperactivity Disorder (ADHD) symptoms. At age 11.5 (SD=0.69) and 16.3 (SD=0.69), parents reported on twins' ADHD symptoms via the Strength and Difficulties Questionnaire (Goodman, 1997) hyperactivity subscale (three-point Likert scale) and the Conners' rating scales (CPRS-R; four-point Likert scale) (Conners, 2003) on hyperactivity and inattention. Although self-report ratings were available, it has been shown that informant-based ratings are more reflective of objective measures of ADHD symptoms (Rietz et al., 2016). A composite score was created as the arithmetic mean of the sex and age z-standardized scales. Where ratings were available at one assessment only, this value was used to maximise sample size, leading to a sample size of 2,469 twin pairs.

Schizophrenia symptoms. At age 22.7 (SD=0.85), paranoia and hallucinations were assessed through self-reported ratings on the Specific Psychotic Experiences Questionnaire (SPEQ; six point Likert scale) (Bell, Halligan, & Ellis, 2006; Fenigstein & Vanable, 1992), and parent-

reported negative symptoms using the Scale for the Assessment of Negative Symptoms (SANS; four point Likert scale) (Andreasen, 1984). Data were available for 1,140 twin pairs.

Family socio-economic status (SES). This measure was calculated as the mean of the z-standardized maternal age at birth of the first child, maternal and paternal highest education level (coded from 1 = "no qualifications" to 8 = "postgraduate qualifications"), and maternal and paternal occupation (coded from 1 = "Other Occupations – dockers, porters, labourers,..." to 9 = "Managers and Administrators"). These measures were assessed at first contact at age 1.8 (SD=1.13). Data were available for 2,962 twin pairs.

Measures were selected based on largest sample sizes available, and ages at phenotype assessment matching most closely the ages of GWA study samples to maximise predictive power. None of the measures were significantly associated with birth order, but most showed sex and age differences (see Supplementary Table S1) and were therefore adjusted for these effects using the regression method, and z-standardised residuals (mean=0, SD=1) were used for all subsequent analyses.

Genotypic data

Two different genotyping platforms were used because genotyping was undertaken in two separate waves, five years apart. AffymetrixGeneChip 6.0 SNP arrays were used to genotype 3,665 individuals. Additionally, 8,122 individuals (including 3,607 dizygotic co-twin samples) were genotyped on Illumina HumanOmniExpressExome-8v1.2 arrays. After quality control, 635,269 SNPs remained for AffymetrixGeneChip 6.0 genotypes, and 559,772 SNPs for HumanOmniExpressExome genotypes.

Genotypes from the two platforms were separately phased and imputed into the Haplotype Reference Consortium (release 1.1) through the Sanger Imputation Service (McCarthy et al., 2016) before merging. Genotypes from a total of 10,346 samples (including 3,320 dizygotic twin pairs and 7,026 unrelated individuals) passed quality control, including 3,057 individuals genotyped on Affymetrix and 7,289 individuals genotyped on Illumina. The identity-by-descent (IBD) between individuals was < 0.05 for 99.5% in the sample excluding the DZ co-twins (range = 0.00 – 0.12), and ranged between 0.36 and 0.62 for the DZ twin pairs (mean=0.49). The final data contained 7,363,646 genotyped or well imputed SNPs (for full

genotype processing and quality control details, see(Selzam et al., 2018)). To ease high computational demands of the software that generates polygenic scores, we further excluded SNPs with info <1, leaving 515,000 SNPs for analysis.

We performed principal component analysis on a subset of 39,353 common (MAF > 5%), perfectly imputed (info = 1) autosomal SNPs, after stringent pruning to remove markers in linkage disequilibrium ($r^2 > 0.1$) and excluding high linkage disequilibrium genomic regions to ensure that only genome-wide effects were detected.

Polygenic scores

We calculated polygenic scores, which are the SNP effect size weighted sums of the number of trait-associated alleles, based on summary statistics for the largest GWA studies available for key developmental outcomes, including height (Yengo et al., 2018), body mass index (BMI) (Yengo et al., 2018), self-rated health (McInnes et al., 2018), intelligence (Savage et al., 2018), educational attainment (Lee et al., 2018), neuroticism (Luciano et al., 2018), ADHD (Demontis et al., 2019), and schizophrenia (Pardiñas et al., 2018). These GWA studies were selected because their respective GPS yield the highest predictive accuracy within their trait category (details about the studies, reported SNP heritabilities and GPS predictions can be found in Supplementary Table S2). To calculate the polygenic scores, we used the software LDpred (Vilhjalmsson et al., 2015) which re-weights the SNP effect sizes based on a prior on the effect size and the LD in the sample. Here, we applied a prior on the fraction of causal markers of 1 for all analyses, based on the assumption that all genetic markers contribute to trait development (see Supplementary Methods S1 for details on polygenic score calculation). All polygenic scores were statistically adjusted for the first ten principal components, chip and plate using the regression method, and were z-standardized (mean=0, SD=1).

Statistical Analysis

Mixed-effects modelling

We applied a random intercept mixed-effects model on DZ data, including two fixed effects to separate the total effect between the polygenic score predictor and the outcome into within- and between-family effects (Carlin, Gurrin, Sterne, Morley, & Dwyer, 2005):

$$Y_{ij} = \alpha_0 + \beta_W\left(GPS_{ij} - \overline{GPS_j}\right) + \beta_B\overline{GPS_j} + \gamma_j + \varepsilon_{ij}, (1)$$

where Y denotes the outcome and GPS the polygenic score, $i = \{1,2\}$ corresponds to the individual twins that are clustered within family j, and $\overline{GPS}$ refers to the mean GPS value in family j. The ith value represents birth order, where twin 1 is the elder twin. The notation α_0 represents the intercept and γ_j the random effect with $\gamma_j \sim N(0, \sigma_\gamma^2)$, which corresponds to a change in the intercept for both twins in family j, and ε_{ij} with $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$, which denotes the independent random error for each individual i in family j. The between-family effect β_B represents the expected change in the outcome Y given a one unit change in the family GPS average, and the within-family effect β_W represents the expected change given a one unit change in the difference between the individual GPS and the family average GPS. By including both β_W and β_B in the same model, the individual estimates are adjusted for, and independent of, the effect of the other estimate. The random effect term σ_γ^2, which estimates the difference between each group intercept γ_j and the overall intercept α_0, accounts for the residual structure in the data corresponding to all unaccounted familial factors (both genetic and environmental) that contribute to the trait similarity of the twins (Carlin et al., 2005; Genser, Teles, Barreto, & Fischer, 2015).

The use of a mixed-effects model is only justified if co-twins within a family correlate in the outcome, which can be estimated through the Intraclass Correlation (ICC) coefficient. The ICC is the ratio of the between-family (i.e. random intercept) variance over the total variance and is an estimate of how much of the total variation in the outcome is accounted for by family:

$$Cor\left(Y_{1j}, Y_{2j}\right) = \frac{\sigma_\gamma^2}{(\sigma_\gamma^2 + \sigma_\varepsilon^2)}, (2)$$

where σ_γ^2 is the covariance between the family variable, in this case family ID, and the outcome, and σ_ε^2 indicates the residual variance capturing within twin pair differences. The

total effect of the relationship between GPS and outcome is the ICC weighted sum of the within- and between-family effects (Genser et al., 2015):

$$Total\ effect = \beta_W(1 - ICC) + \beta_B\, ICC. \quad (3)$$

It follows from (3) that the total effect ranges between β_W and β_B. If the relationship between GPS and outcome is mostly due to individual-level variation, the ICC approximates 0 and the total effect will be close to β_W. In contrast, if the association is mostly due to family effects, the ICC approximates 1 and the total effect will be close to β_B (Genser et al., 2015). To calculate the total effect, we used ICC estimates adjusted for the fixed effects described in equation (1).

Performing a regression corresponding to equation (1), we estimated the β_W and β_B parameters using each of the eight polygenic scores in turn as predictors of each of the eight measured outcomes. To estimate potential SES effects, we repeated these analyses including the SES composite as a covariate in the model (these latter analyses were not preregistered). For the fixed effects, we calculated 95% bootstrap percentile intervals. These were based on 10,000 bootstrap samples with random resampling of DZ twin pairs with replacement.

To empirically test the statistical difference between β_W and β_B, we divided the difference between the fixed effect coefficients by the standard deviation of the sampling distribution of the estimate differences (Clogg, Petkova, & Haritou, 1995; Paternoster, Brame, Mazerolle, & Piquero, 1998). We also applied this approach to statistically test the significance of the difference between the β_B coefficients before and after the inclusion of family SES in model (1). To evaluate the effect size change between the coefficients, we calculated the beta differences with 95% bootstrap percentile intervals, as well as the percentage change (e.g. $(\beta_B - \beta_W) / \beta_B$).

Quantile analysis of within-DZ pair differences

To illustrate the extent to which within-DZ pair GPS differences result in differences in developmental outcomes, we performed quantile analysis. Firstly, we generated twin-GPS difference scores by subtracting the twin 2 score from the twin 1 score, and then split this variable into ten equal quantiles based on absolute GPS differences, ranging from the lowest

to the highest GPS differences. Birth order did not explain any statistically significant amount of variance (Supplementary Table S1), therefore no randomisation of twin order was required. We tested mean differences in outcome variables between individuals in the lowest and highest decile. We performed quantile analysis on variables with scales that are easily interpretable: that is, BMI, height, intelligence and educational achievement. For this purpose, the z-standardised and cleaned variables were transformed back to their original scale, and intelligence values were scaled to have a mean of 100 and a standard deviation of 15.

Multiple testing correction

Multiple testing correction of the significance threshold was performed using the Benjamini Hochberg false discovery rate (FDR) adjustment (Benjamini & Hochberg, 1995). In contrast to more conservative corrections, this method has higher statistical power to detect true positives while controlling for false positives. Based on an α threshold of 0.05, the corrected α in this study was 0.01, defined as the maximum raw *p*-value that is smaller or equal than the FDR critical value ($p_{raw} \leq \frac{rank\ of\ p_{raw}}{total\ number\ of\ p_{raw}\ values} \times \alpha$).

Sensitivity analyses

We performed additional, non-pre-registered sensitivity analyses to evaluate the robustness of our findings. Mixed-effects models were run separately for same-sex and for opposite-sex twin pairs (for twin pair *N*, see Supplementary Table S3), as well as for twin pairs where both twins were genotyped on the OEE chip, and twin pairs where one twin was genotyped on OEE and the other twin genotyped on Affymetrix (for twin pair *N*, see Supplementary Table S4). Analyses were also performed using GPSs that were constructed applying a prior based on a causal fraction of 0.1.

To control for any unaccounted relatedness between families, we estimated the fixed effects including a SNP-kinship matrix as random effect. Here, equation (3) becomes $y = \alpha + \beta_W\left(GPS_{ij} - \overline{GPS_j}\right) + \beta_B\overline{GPS_j} + g + \varepsilon$, where g is the random effect with $g \sim N(0, A\sigma_g^2)$ and A being a genetic relationship matrix between individuals. A pairwise genetic relationship between individual m and n is estimated as $A_{mn} = 1/N \sum_{i=1}^{N}(x_{im} -$

$2p_i)(x_{in} - 2p_i)/2p_i(1 - p_i)$, where N is the number of SNPs, x_{im} is the number of copies of the reference allele for the i^{th} SNP of the m^{th} individual and p_i is the reference allele frequency. These analyses were performed using the GCTA software (version 1.90.0) (Yang, Lee, Goddard, & Visscher, 2011).

Due to the large study population of the UK Biobank, there may be relatedness between this sample and the UK target sample TEDS. The UK Biobank sample was included in the GWA meta-analysis of height, BMI and educational attainment, and relatedness between discovery and target sample could lead to GPS prediction estimate inflation in the target sample (Choi, Mak, & O'Reilly, 2018; Wray et al., 2013). We therefore calculated an additional set of height and BMI polygenic scores based on GWA meta-analyses published before UK Biobank data became available (Locke et al., 2015; Wood et al., 2014). We also calculated an additional GPS for educational attainment based on a GWA analysis that had all British cohorts removed (Lee et al., 2018). While this rules out discovery and target sample relatedness, it also controls for effect size inflation due to population stratification.

Results

Phenotypic resemblance between DZ twins within a family varied across traits, with Pearson's correlation coefficients ranging from 0.10 – 0.59 (Supplementary Figure S1, and Supplementary Table S5 for ICCs). Twins were least alike in their neuroticism levels and self-rated health, and most alike in their height, IQ and educational achievement. Within-twin pair polygenic score correlations were close to expectations (range $r = 0.49 – 0.57$), as the expected shared additive genetic variance between siblings is 50% of the total additive genetic variance based on quantitative genetic theory (Fisher, 1918). Given the 95% confidence intervals of the within-twin pair correlations (Supplementary Figure S1), there was a significant difference from the expected correlation coefficient of 0.50 for the self-rated health GPS ($r = 0.53$), the IQ GPS ($r = 0.54$), and the educational attainment GPS ($r = 0.57$), indicating assortative mating for these traits.

Within-family polygenic score predictions

Figure 1A depicts the within- and between-family polygenic score prediction estimates of the eight outcomes from the mixed-effects model analyses. Within-family target-trait predictions were statistically significant for height, BMI, intelligence, educational achievement and ADHD symptoms, indicating that polygenic variation within twin pairs was related to these outcome differences. Specifically, phenotypic differences in height were significantly positively correlated with height GPS twin differences ($\beta = 0.41$, $p = 5.72e^{-53}$) and differences in BMI were significantly correlated with BMI GPS differences ($\beta = 0.30$, $p = 1.76e^{-21}$) such that twins with a higher height GPS and BMI GPS were taller and heavier than their co-twin, respectively. IQ GPS differences predicted intelligence differences ($\beta = 0.14$, $p = 1.32e^{-6}$) and EA GPS differences were significantly associated with GCSE grade differences ($\beta = 0.21$, $p = 2.22e^{-26}$), indicating that those twins with a higher GPS also scored higher on intelligence measures and in their GCSE tests than their co-twin. For behaviour problems, twins with higher ADHD GPS had higher phenotypic ADHD symptoms than their co-twins ($\beta = 0.12$, $p = 1.50e^{-7}$).

We also investigated cross-trait relationships (Figure 1). For example, self-rated health GPS differences were negatively correlated with differences in BMI, such that twins with a higher

self-rated health GPS had a lower BMI ($\beta = -0.13$, $p = 3.56e^{-5}$). EA GPS differences significantly related to phenotypic intelligence differences ($\beta = 0.13$, $p = 2.15e^{-5}$), and IQ GPS predicted GCSE grade differences ($\beta = 0.20$, $p = 7.24e^{-25}$), suggesting that those with higher GPSs also had higher IQ and GCSE grades than their co-twin. GCSE grade differences were also negatively predicted by ADHD GPS twin differences ($\beta = -0.07$, $p = 2.20e^{-4}$), indicating that twins with a higher ADHD GPS obtain lower GCSE results. Notably, IQ GPS differences ($\beta = -0.12$, $p = 6.38e^{-7}$) and EA GPS differences ($\beta = -0.14$, $p = 3.09e^{-8}$) were just as predictive of ADHD symptoms as the ADHD GPS itself, and the direction of effect sizes indicates that the twin with a higher GPS had lower ADHD symptoms than their co-twin (all prediction estimates and total effects are presented in Supplementary Table S6).

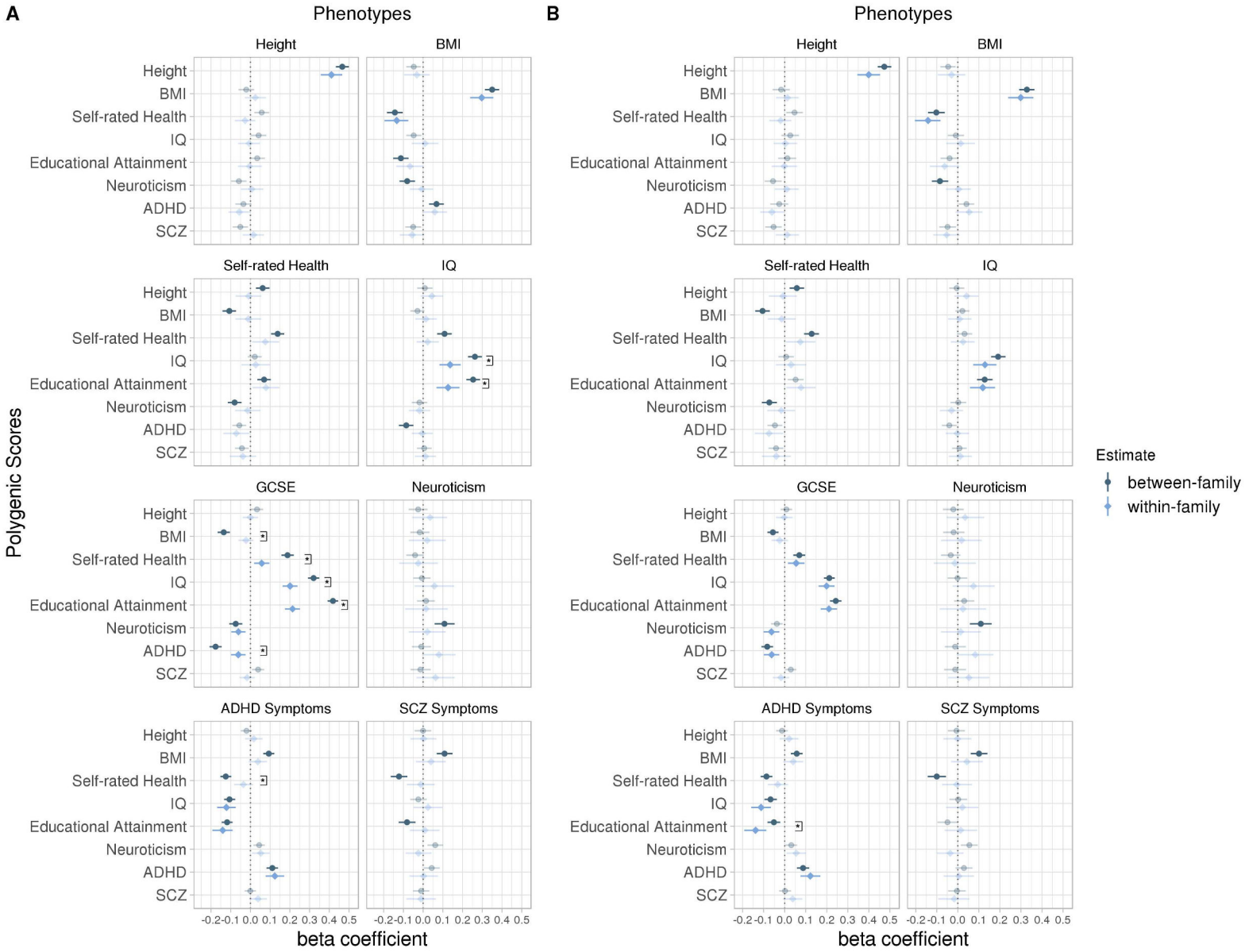

A
Phenotypes
Height
BMI
Self-rated Health
IQ
Educational Attainment
Neuroticism
ADHD
SCZ
Polygenic Scores
-0.2 -0.1 0.0 0.1 0.2 0.3 0.4 0.5
beta coefficient
B
Phenotypes
Estimate
between-family
within-family

Figure 1. Within- and between-family prediction estimates of eight outcomes using eight genome-wide polygenic scores. Panel A shows findings before statistical correction for family socio-economic status (SES), and panel B shows findings after correction for SES. The genome-wide polygenic scores (GPS) are presented on the y-axis, predicting each of the eight phenotypic traits. Error bars are 95% bootstrap percentile intervals based on 10,000 bootstrap samples (random resampling of DZ twin pairs with replacement). Opaque estimates indicate statistical significance at the false discovery rate corrected threshold of $p < 0.01$. Brackets indicate a significant difference between within- and between-family prediction estimates. Significant differences are only shown where at least one of the estimates is statistically significant at the false discovery rate corrected threshold of $p < 0.01$ (for all prediction estimates and p-values, see Supplementary Tables S6 and S7). The dotted line represents a beta coefficient of zero. BMI = Body Mass Index; IQ = Intelligence; GCSE = General Certificate of Secondary Education (educational achievement); ADHD = Attention-Deficit/Hyperactivity Disorder; SCZ = Schizophrenia.

Comparing within-family and between-family polygenic score prediction

By simultaneously and independently estimating within- and between-family GPS predictions, it was possible to compare these estimates. Between-family estimates (Figure 1A) are mostly consistent with GPS correlations reported for unrelated individuals (Supplementary Table S2). Figure 1 also shows that between-family associations are generally greater than within-family associations. Significant associations were found for 46.9% of the between-family associations and only 20.3% for within-family associations. On average, magnitudes of within-family associations were almost half (44.1% reduction) that compared to significant between-family estimates (for all prediction estimates, beta difference values and their 95% confidence intervals, and significance of differences see Supplementary Table S6).

Notably, significant differences in associations within and between families for polygenic scores predicting their target traits were almost exclusively found for IQ and educational achievement (Figure 1A). The within-family prediction was significantly lower than between-family prediction for both IQ ($p = 6.27e^{-4}$, $\Delta = 48.0\%$) and GCSE grades ($p = 8.45e^{-14}$, $\Delta = 48.9\%$). Despite not reaching statistical significance, we also observed attenuation of the within-family prediction relative to the between-family prediction for height ($\Delta = 11.8\%$), BMI ($\Delta = 15.1\%$), self-rated health ($\Delta = 45.2\%$) and neuroticism ($\Delta = 80.4\%$).

Also, for cross-trait associations, differences in within- and between-family polygenic score predictions were most pronounced for IQ and educational achievement. For IQ, there were significant differences for the EA GPS ($p = 7.57e^{-4}$, $\Delta = 50.1\%$). For educational achievement, there were significant differences for the BMI GPS ($p = 8.10e^{-5}$, $\Delta = 83.3\%$), the self-rated health GPS ($p = 4.60e^{-6}$, $\Delta = 69.5\%$), the IQ GPS ($p = 1.79e^{-5}$, $\Delta = 37.2\%$), and the ADHD GPS ($p = 4.95e^{-5}$, $\Delta = 65.4\%$). In addition, there was a significant difference in within- and between-family prediction for the self-rated health GPS ($p = 4.00e^{-3}$, $\Delta = 71.7\%$) predicting ADHD symptoms. Although not significant, effect size attenuations were also sizeable for other cross-trait predictions, such as for the neuroticism GPS predicting BMI ($\Delta = 91.2\%$), or the self-rated health GPS predicting schizophrenia symptoms ($\Delta = 90.1\%$) (Supplementary Table S6). However for these comparisons, between-family coefficients may not be as reliable as the

between-family coefficients that a showed significant difference to their within-family estimate, as estimates were considerably smaller to begin with.

The finding that polygenic score prediction estimates of our measured traits are substantially smaller within families suggests that the corresponding between-family associations are mediated by some combination of family-specific (i.e. shared family) effects, population stratification and potentially assortative mating. Family SES, which is the same for members of a family, is a predictor not only of educational achievement and IQ, but also physical and mental health outcomes. Therefore, we repeated our analyses including family SES as a covariate in the model to interrogate its role in between-family GPS prediction. As noted above, this analysis was not pre-registered. As shown in Figure 1B, between-family predictions were greatly reduced and magnitudes approached those of within-family prediction estimates, which did not change (because any shared family effects are already controlled for in within-family estimates; for all prediction estimates, beta difference values and their 95% confidence intervals, and significance of differences, see Supplementary Table S7).

Formal testing of the between-family estimate differences before and after correcting for SES indicated significant differences only for cognitive traits (Supplementary Table S8, Supplementary Figure S2). For example, there was an average attenuation of 60.9% across the within- and between-family comparisons for the GPSs that showed a statistically significant difference in their prediction of GCSE grades, which was reduced to 25.8% after accounting for SES. Although this is a substantial attenuation, these findings show that family SES does not account for all of the observed differences.

We performed additional contrasts, controlling for the SES components parental education and parental occupation separately in an attempt to identify more specific potential sources of prGE. For GCSE grades and IQ, between-family beta coefficients showed greater attenuation when controlling for parental education in comparison with parental occupation (Supplementary Tables S9 and S10, Supplementary Figure S3). However, only for the educational attainment GPS predicting GCSE grades was the difference between the attenuation due to parental education (Δ = 18.9%) and parental occupation (Δ = 37.7%) statistically significant (p = 6.40e^{-3}) (Supplementary Table S11), indicating that parental education may present a stronger prGE effect.

As a further set of analyses, we applied a multiple regression approach to predict family SES using the within- and between-family estimates of the eight GPSs. Family SES acts as a control trait as there should be no direct genetic effects from the offspring to family SES, as indicated by the within-family effect. Results confirmed that all within-family beta coefficients were zero, while between-family estimates were related to family SES (Supplementary Table S12).

Sensitivity analyses (not pre-registered) were performed by repeating all analyses separately for same-sex and opposite-sex twins (Supplementary Tables S3 and S13 to S16, and Supplementary Figures S4 and S5), and for twin pairs grouped by genotyping chip (Supplementary Tables S4 and S17 to S20, and Supplementary Figures S6 and S7). In addition, we estimated the fixed effects using a SNP-kinship matrix as a random effect to control for any unaccounted between-family relatedness (Supplementary Table S21 and Supplementary Figure S8). For the different sets of sensitivity analyses described, no substantial deviations from the results using the combined sample were found.

We also repeated analyses using GPSs that were calculated based on a fraction of causal markers of 0.1 (Supplementary Tables S22 and S23, and Supplementary Figure S9), and using GPSs that had the UK Biobank sample (height; BMI) or all British samples (educational attainment) removed at the GWA analysis stage (Supplementary Table S24 and Supplementary Figure S10). Although prediction estimates were smaller in some cases likely due to reduced power, the pattern of within- versus between-family effect size changes remained unchanged as indicated by the mostly overlapping 95% confidence intervals of the beta difference values.

Quantile Analysis

To illustrate within-family differences further, quantile analysis demonstrated how within-family polygenic score differences related to differences in height, BMI, intelligence and GCSE grades (Figure 2). There was an 8.7cm height mean difference ($p = 1.28e^{-11}$) between the lowest absolute difference decile versus the highest difference decile. For BMI, the difference was 2.9 BMI points ($p = 8.33e^{-6}$) between the lowest and the highest absolute GPS difference deciles. Mean GCSE grade differences (0.40) were also statistically significant ($p =$

7.13e^{-5}) when comparing the lowest and the highest absolute GPS difference deciles. In contrast, intelligence point differences (1.9 points) were not statistically different (p = 0.26) between the lowest and the highest absolute GPS difference quantiles (for trait and GPS means at each difference decile see Supplementary Table S25).

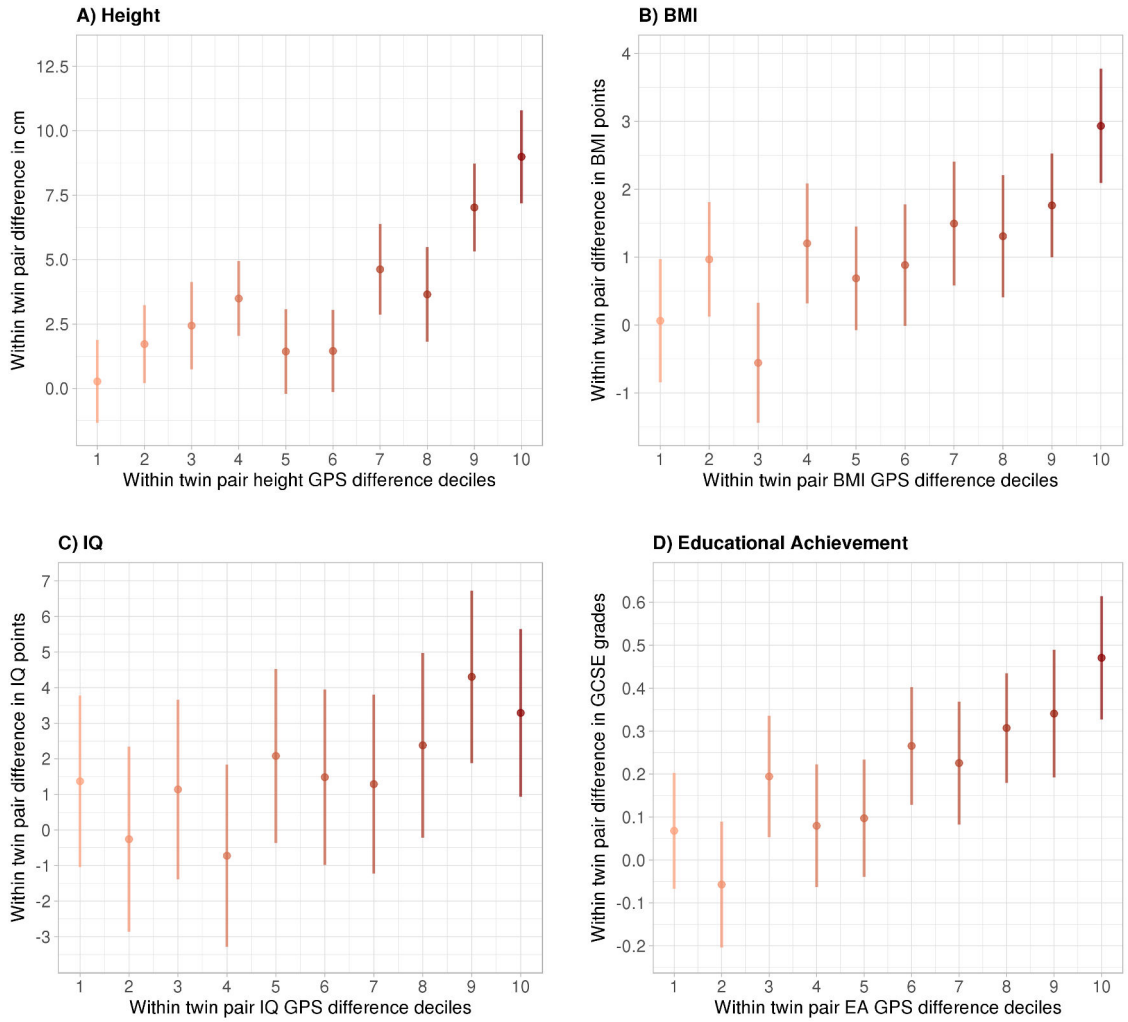


Figure 2. The relationship between absolute dizygotic (DZ) twin pair polygenic score decile differences and trait outcome differences. Lower deciles represent small absolute genome-wide polygenic score (GPS) differences and higher deciles represent large GPS differences between DZ co-twins. Error bars indicate 95% confidence intervals. Each GPS decile included the following numbers of twin pairs: Height = 146; BMI = 135; IQ = 157; GCSE = 236. Regression through origin analysis (fixed intercept of zero) using the continuous GPS difference values to predict outcome differences were significant for height (B = 4.42, p = 3.73e-53, R^2 = 0.148), BMI (B = 1.34, p = 1.73e-21, R^2 = 0.064), IQ (B = 2.1, p = 4.53e-7, R^2 = 0.015), and GCSE grades (B = 0.26, p = 3.04e-26, R^2 = 0.046).

Discussion

Polygenic score prediction of complex traits is now a common approach in genomics research, but the potential pathways by which polygenic score variation predicts phenotypic variation remain largely unexplored. In this study, we contrasted within- and between-family polygenic prediction estimates to quantify the extent to which environmentally-mediated genetic effects (i.e. passive genotype-environment correlation) are picked up in polygenic score analyses. By systematically performing target- and cross-trait analyses across eight life outcomes using eight corresponding GPS, we found evidence that prGE might be a mechanism explaining a considerable proportion of the GPS prediction in cognitive traits (intelligence and educational achievement), but not as much for non-cognitive traits. We also found that for between-family GPS predictions of cognitive traits – but, again, not as much for other traits – family SES is likely to be the major source of prGE.

For the prediction of IQ and educational achievement, within-family estimates were on average 60% smaller than between-family estimates. The within- versus between-family attenuation for the EA GPS prediction was 49%, which is close to the 40% estimate in GWA study effect sizes for years of education (Lee et al., 2018), and the 55% estimate using the same EA GPS in a different target sample (Trejo & Domingue, 2019). These findings highlight the influence of prGE in the development of IQ and educational achievement, and demonstrate the extent to which between-family GPS prediction may be partly driven by prGE effects. Results from our study are also in line with adoption studies showing evidence of between-family prGE in that correlations between home environment and children's IQ is twice as great in non-adoptive families than in adoptive families (Plomin, 1994). Our findings are compatible with recent research on *genetic nurture*, using non-transmitted alleles from parental genotypes to assess prGE (Bates et al., 2018; Kong et al., 2018) in terms of GPS target trait prediction of educational achievement and anthropometric traits. Our findings also extend to cross-trait associations using a wide range of GPSs. Contrary to our prediction that within- and between-family EA GPS associations would be significantly different across many associated life outcomes, results from cross-trait analysis suggest that within- and between-family predictions were only significantly different across a range of GPS for the prediction of cognitive outcomes.

A possible explanation for these results is that IQ and educational achievement show more shared environmental influences (24% and 27%, respectively) relative to other traits used in this study such as height (10%), BMI (10%), ADHD (2%), or schizophrenia (0%), as estimated through a large twin study meta-analysis (Polderman et al., 2015). The type of rGE that we assessed in this study – defined as the exposure to a family environment that is correlated with both parental and offspring genotypes, and which contributes to sibling similarity in their outcomes – is absorbed by the shared environment variance component ('C') in classical twin analyses (Rijsdijk, 2002). Therefore, it may be more likely that genetic effects related to cognitive traits as estimated through GWA studies partly contain prGE effects – in contrast to other traits tested in our study – because the shared environmental component is larger to begin with for cognitive traits. In TEDS, the C component for the same IQ and educational achievement measures used in this study were estimated around 35% (Kovas et al., 2013) and 29% (Shakeshaft et al., 2013), respectively.

As known from the existing literature, family SES is strongly genetically correlated with offspring cognitive traits (Hill et al., 2016; Krapohl & Plomin, 2016; Trzaskowski et al., 2014), rendering it a likely source of prGE. Indeed, our results showed that between-family effects were considerably more similar in magnitude to within-family effects when holding SES constant, suggesting that SES is a source of the majority of the within-between discrepancy, rather than residual population stratification or assortative mating. When controlling for parental education and parental occupation separately, we found that between-family effect sizes were closer to within-family coefficients for parental education than for occupation. However, this difference was only significant for the educational attainment GPS predicting GCSE grades, suggesting that parental education is likely a stronger source of prGE than parental occupation influencing offspring educational achievement.

Despite the sizeable attenuations after controlling for family SES, we still observed some effect size differences when comparing within- and between-family coefficients. For example, there was still a 32.5% difference for the IQ GPS predicting IQ and a 13.4% difference for the educational attainment GPS predicting GCSE scores. The within-twin pair correlations for these GPS indicated assortative mating, which could explain some of this remaining discrepancy. Indeed, previous research on genetic nurture indicated that a small proportion of the direct genetic effect of the educational attainment polygenic score predicting educational attainment captures assortative mating-related effects (Kong et al.,

2018). The same research also showed genetic nurture effects between siblings using the educational attainment GPS (Kong et al., 2018). Such effects may further contribute to the within-family effect attenuation, potentially accounting for some of the residual difference after controlling for family SES.

The results showed that more distantly-related GPS captured considerable prGE effects in cross-trait GPS predictions of cognitive traits. For instance, within-family effect sizes for the ADHD GPS predicting educational achievement were significantly smaller (65% reduction), in contrast to the ADHD GPS predicting ADHD symptoms, where no difference was detected. This suggests that the GWAS for ADHD captures genetic variation that is correlated with aspects of the family environment that contribute to the co-development of ADHD symptoms and educational achievement, although it is unclear why these effects do not appear to contribute to the development of ADHD symptoms themselves.

It is important to go beyond GPS predictions of traits in unrelated individuals to consider prGE mechanisms by comparing within- and between-family predictions in order to explain the sources of predictions in polygenic score analysis. However, finding between-family prGE does not diminish the usefulness of GPS predictions for cognitive traits in unrelated individuals, because these prGE effects help maximise the prediction of trait variance. Although within-family genetic effects do not include prGE effects due to between-family factors such as SES, within-family genetic effects are not free of *all* kinds of rGE, as demonstrated by twin studies showing that correlations between putative measures of the environment and children's specific outcomes are genetically influenced (Plomin, 1994). Within-family GPS prediction estimates can be interpreted as direct genetic effects in the sense that they stem from the individual level not the family level. Children select, modify and create experiences (active rGE), or evoke responses in their environment (evocative rGE) that are correlated with their genetic propensities. Therefore, within-family genetic differences can relate to trait differences through active or evocative rGE pathways, but are free of any passive rGE effects.

Implications

The results from this study have three important implications for the interpretation of the existing polygenic score literature, as well as for future genetic research. First, the finding

that between-family predictions pick up effects due to prGE mostly and substantially in cognitive traits is informative for causal inference studies that use designs such as Mendelian Randomisation (Smith & Ebrahim, 2005; Smith & Hemani, 2014). Here, a genetic instrument that is related to a predictor (in form of a single genetic marker or GPS) is used to assess the causal relationship between the predictor and an outcome. At a population level, genotypes are not inherited randomly: individuals with particular genotypes are not born into environmental conditions at chance. If family environment is associated with the genetic instrument as well as the predictor and the outcome, this opens a backdoor path whereby predictor and outcome are related through the prGE mechanisms (Pingault et al., 2018). This could lead to an assumption violation, therefore biasing causal inference in between-family analysis. Only in a within-family design is it ensured that Mendelian Randomisation meets its assumptions because transmission of alleles is randomised at meiosis within families, and because prGE effects due to shared environment are held constant (Brumpton et al., 2019; Pingault et al., 2018; Smith, 2007; Smith & Ebrahim, 2003). Although genetic data for siblings are often not available, our results provide a useful guideline for the GPS-outcome combinations that are unlikely to suffer from this assumption violation when applying designs such as Mendelian Randomisation to unrelated samples. For example, our results indicate that caution should be warranted due to prGE mechanisms if applying Mendelian Randomisation to cognitive traits, even if family SES is included as a confounder in the analyses as confounding effects might not be captured perfectly. On the contrary, other traits such as BMI and ADHD (with the possible exclusion of the self-rated health GPS) should be less problematic, because within- and between-family effect sizes match closely, ruling out potential confounding due to prGE.

Second, our results provide evidence that location-related population stratification is not a large bias in GPS prediction of complex traits when controlling for genetic principal components in samples from White European backgrounds. As it has been shown that the GPS prediction of height is affected by population stratification (Berg et al., 2019), we also find an attenuation of around 12% of the within-family coefficient, which is by necessity free of population stratification since stratification is constant within a family. When we performed our analyses using a GPS for height based on a discovery sample that did not include UK Biobank, the attenuation decreased to 5%. This may indicate that the inclusion of a large discovery sample genetically similar to the target sample could have resulted in a GPS that is more strongly confounded by population stratification – although it is

noteworthy that the 95% confidence intervals of the beta difference values overlap for the two height GPSs. For those traits where within- and between-family estimate differences were large and significant, differences were greatly reduced after accounting for SES, indicating that SES was the main source of the discrepancy, as opposed to location-related population stratification. Our additional analyses using a GPS based on GWA analysis that had all British samples removed did not show less attenuation, which would be expected if population stratification strongly influenced GPS prediction.

Third, our study illustrates the usefulness of obtaining genotypic data on family members, since it makes it possible to identify mechanisms of polygenic prediction. Our results demonstrate that by analysing DZ co-twins' genetic data jointly, prGE mechanisms due to shared environment (and in this case associated with SES) can be revealed.

Limitations

Although we present the most comprehensive within- and between- family comparison of GPS prediction to date, there are limitations to this study. The GWA studies used to generate the eight GPS for this study had different statistical power to discover genetic effect sizes due to sample size variations and different underlying genetic architectures of the GWA study traits. As a result, each of the eight GPSs were differently powered to detect target- and cross-trait associations, making it difficult to draw direct comparisons across the within- and between-family prediction effect sizes. Lack of power may also lead to an inability to detect small prGE effects that would become visible with (i) more powerful GPS and (ii) the availability of larger DZ twin pair samples. However, we detected prGE effects in cross-trait analysis using the ADHD GPS, which is based on the smallest GWAS study sample (~55,000 individuals), indicating that we had sufficient power to detect at least some of the prGE effects.

It is also possible that some important within- and between-family effect differences did not reach statistical significance due to insufficient statistical power. While the effect size differences in cognitive traits are large, it may be that effects due to prGE, population stratification and/or assortative mating are more subtle in other traits. Therefore, our study sample, which ranged between 789 and 2,469 DZ twin pairs, may have not had enough power to establish the statistical significance of small effect size differences. Notably, GPS

predictions were generally small where no significant difference was found between large within- versus between-family effect size attenuations. With the availability of more powerful GPS in the future, it may be possible to detect such differences statistically.

Another limitation was that we did not have parental genotypes available to directly test the influence of non-transmitted parental alleles on offspring outcomes (genetic nurture) (Kong et al., 2018). Although the within-family design used in this study accounts for the effects of both transmitted and non-transmitted parental alleles on offspring outcomes, it is not possible to disentangle these two sources of prGE. Future studies would benefit from incorporating parental and sibling genotypes to disentangle the prGE effects through the joint analysis of parental and sibling genotypes, which will shed light on how both non-transmitted parental and non-co-inherited sibling alleles contribute to trait development.

Conclusion

This study provided strong evidence for prGE mechanisms in polygenic score prediction mainly for cognitive traits across a range of different polygenic scores. The implications of these findings for future studies depend on their aims. If maximising trait prediction is the goal, the use of unrelated samples is valid even in the presence of prGE effects because these influences are informative nonetheless. However, if the goal is causal inference and explanation, a within-family genetic design is recommended to avoid prGE-related confounding. The increasing availability of genotypic data in relatives will become a crucial element in genetics research, allowing researchers to disentangle the mechanisms of polygenic prediction of complex human traits.

Online Resources

OSF pre-registration link

https://osf.io/eq8ga/?view_only=768f42366d134ebeb50f5999763c3fce

References

Allegrini, A., Selzam, S., Rimfeld, K., Stumm, von, S., Pingault, J.-B., & Plomin, R. (2019). Genomic prediction of cognitive traits in childhood and adolescence. *Molecular Psychiatry*. http://doi.org/https://doi.org/10.1101/418210

Andreasen, N. C. (1984). Scale for the Assessment of Negative Symptoms.

Bates, T. C., Maher, B. S., Medland, S. E., McAloney, K., Wright, M. J., Hansell, N. K., et al. (2018). The Nature of Nurture: Using a Virtual-Parent Design to Test Parenting Effects on Children's Educational Attainment in Genotyped Families. *Twin Research and Human Genetics*, *21*(2), 73–83. http://doi.org/10.1017/thg.2018.11

Bell, V., Halligan, P. W., & Ellis, H. D. (2006). The Cardiff Anomalous Perceptions Scale (CAPS): a new validated measure of anomalous perceptual experience. *Schizophrenia Bulletin*, *32*(2), 366–377. http://doi.org/10.1093/schbul/sbj014

Belsky, D. W., & Harden, K. P. (2019). Phenotypic Annotation: Using Polygenic Scores to Translate Discoveries From Genome-Wide Association Studies From the Top Down. *Current Directions in Psychological Science*, 0963721418807729. http://doi.org/10.1177/0963721418807729

Belsky, D. W., Domingue, B. W., Wedow, R., Arseneault, L., Boardman, J. D., Caspi, A., et al. (2018). Genetic analysis of social-class mobility in five longitudinal studies. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(31), E7275–E7284. http://doi.org/10.1073/pnas.1801238115

Belsky, D. W., Moffitt, T. E., Corcoran, D. L., Domingue, B., Harrington, H., Hogan, S., et al. (2016). The Genetics of Success. *Psychological Science*, *27*(7), 957–972. http://doi.org/10.1177/0956797616643070

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, *57*(1), 289–300.

Benyamin, B., Visscher, P. M., & McRae, A. F. (2009). Family-based genome-wide association studies. *Pharmacogenomics*, *10*(2), 181–190. http://doi.org/10.2217/14622416.10.2.181

Berg, J. J., Harpak, A., Sinnott-Armstrong, N., Joergensen, A. M., Mostafavi, H., Field, Y., et al. (2019). Reduced signal for polygenic adaptation of height in UK Biobank. *eLife*, *8*, 68. http://doi.org/10.7554/eLife.39725

Briley, D. A., Domiteaux, M., & Tucker-Drob, E. M. (2014). Achievement-Relevant Personality: Relations with the Big Five and Validation of an Efficient Instrument. *Learning and Individual Differences*, *32*, 26–39. http://doi.org/10.1016/j.lindif.2014.03.010

Brumpton, B., Sanderson, E., Pires Hartwig, F., Harrison, S., Aberge Vie, G., Cho, Y., et al. (2019). Within-family studies for Mendelian randomization: avoiding dynastic, assortative mating, and population stratification biases. *bioRxiv*. http://doi.org/https://doi.org/10.1101/602516

Carlin, J. B., Gurrin, L. C., Sterne, J., Morley, R., & Dwyer, T. (2005). Regression models for twin studies: a critical review. *International Journal of Epidemiology*, *34*(5), 1089–1099. http://doi.org/10.1093/ije/dyi153

Choi, S. W., Mak, T. S. H., & O'Reilly, P. (2018). A guide to performing Polygenic Risk Score analyses. *bioRxiv*, 416545. http://doi.org/10.1101/416545

Clogg, C. C., Petkova, E., & Haritou, A. (1995). Statistical Methods for Comparing Regression Coefficients Between Models. *American Journal of Sociology*, *100*(5), 1261–1293. http://doi.org/10.2307/2782277?refreqid=search-gateway:db6104945ee71e279799239cc06f27ad

Conners, C. K. (2003). Conners' Rating Scales-Revised: Technical Manual (Vol. New York).

De Ridder, K. A. A., Pape, K., Johnsen, R., Holmen, T. L., Westin, S., & Bjorngaard, J. H. (2013). Adolescent Health and High School Dropout: A Prospective Cohort Study of 9000 Norwegian Adolescents (The Young-HUNT). *Plos One*, *8*(9). http://doi.org/10.1371/journal.pone.0074954

de Zeeuw, E. L., van Beijsterveldt, C. E. M., Glasner, T. J., Bartels, M., Ehli, E. A., Davies, G. E., et al. (2014). Polygenic Scores Associated With Educational Attainment in Adults Predict Educational Achievement and ADHD Symptoms in Children. *American Journal of Medical Genetics Part B-Neuropsychiatric Genetics*, *165*(6), 510–520. http://doi.org/10.1002/ajmg.b.32254

Demontis, D., Walters, R. K., Martin, J., Mattheisen, M., Als, T. D., Agerbo, E., et al. (2019). Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nature Genetics*, *51*(1), 63–75. http://doi.org/10.1038/s41588-018-0269-7

Domingue, B. W., Belsky, D. W., Conley, D., Harris, K. M., & Boardman, J. D. (2015). Polygenic Influence on Educational Attainment. *AERA Open*, *1*(3), 1–13. http://doi.org/10.1177/2332858415599972

Fenigstein, A., & Vanable, P. A. (1992). Paranoia and self-consciousness. *Journal of Personality and Social Psychology: Personality Processes and Individual Differences*, *62*(1), 129–138.

Fisher, R. A. (1918). The Correlation between Relatives on the Supposition of Mendelian Inheritance. *Transactions of the Royal Society of Edinburgh*, *52*, 399–433. http://doi.org/10.1017/S0080456800012163

Fletcher, J. M. (2011). The promise and pitfalls of combining genetic and economic research. *Health Economics*, *20*(8), 889–892. http://doi.org/10.1002/hec.1745

Genser, B., Teles, C. A., Barreto, M. L., & Fischer, J. E. (2015). Within- and between-group regression for improving the robustness of causal claims in cross-sectional analysis. *Environmental Health*, *14*(1), 60. http://doi.org/10.1186/s12940-015-0047-2

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A Research Note. *Journal of Child Psychology and Psychiatry*, *38*(5), 581–586. http://doi.org/10.1111/j.1469-7610.1997.tb01545.x

Gratten, J., Wray, N. R., Keller, M. C., & Visscher, P. M. (2014). Large-scale genomics unveils the genetic architecture of psychiatric disorders. *Nature Neuroscience*, *17*(6), 782–790. http://doi.org/10.1038/nn.3708

Hagenaars, S. P., Harris, S. E., Davies, G., Hill, W. D., Liewald, D. C. M., Ritchie, S. J., et al. (2016). Shared genetic aetiology between cognitive functions and physical and mental health in UK Biobank (N=112 151) and 24 GWAS consortia. *Molecular Psychiatry*, *21*(11), 1624–1632. http://doi.org/10.1038/mp.2015.225

Haworth, C. M. A., Davis, O. S. P., & Plomin, R. (2013). Twins Early Development Study (TEDS): A Genetically Sensitive Investigation of Cognitive and Behavioral Development From Childhood to Young Adulthood. *Twin Research and Human Genetics*, *16*(1), 117–125. http://doi.org/10.1017/thg.2012.91

Hill, W. D., Hagenaars, S. P., Marioni, R. E., Harris, S. E., Liewald, D. C. M., Davies, G., et al. (2016). Molecular Genetic Contributions to Social Deprivation and Household Income in UK Biobank. *Current Biology*, *26*(22), 3083–3089. http://doi.org/10.1016/j.cub.2016.09.035

Kaplan, E., Fein, D., Kramer, J., Delis, D., & Morris, R. (1999). WISC-III as a Process Instrument (WISC-III-PI). New York: The Psychological Corporation.

Koellinger, P. D., & Harden, K. P. (2018). Using nature to understand nurture. *Science*, *359*(6374), 386–387. http://doi.org/10.1126/science.aar6429

Kong, A., Thorleifsson, G., Frigge, M. L., Vilhjalmsson, B. J., Young, A. I., Thorgeirsson, T. E., et al. (2018). The nature of nurture: Effects of parental genotypes. *Science*, *359*(6374), 424–428. http://doi.org/10.1126/science.aan6877

Kovas, Y., Voronin, I., Kaydalov, A., Malykh, S. B., Dale, P. S., & Plomin, R. (2013). Literacy and Numeracy Are More Heritable Than Intelligence in Primary School. *Psychological Science*, *24*(10), 2048–2056. http://doi.org/10.1177/0956797613486982

Krapohl, E., & Plomin, R. (2016). Genetic link between family socioeconomic status and children's educational achievement estimated from genome-wide SNPs. *Molecular Psychiatry*, *21*(3), 437–443. http://doi.org/10.1038/mp.2015.2

Krapohl, E., Euesden, J., Zabaneh, D., Pingault, J.-B., Rimfeld, K., Stumm, von, S., et al. (2016). Phenome-wide analysis of genome-wide polygenic scores. *Molecular Psychiatry*, *21*(9), 1188–1193. http://doi.org/10.1038/mp.2015.126

Krapohl, E., Rimfeld, K., Shakeshaft, N. G., Trzaskowski, M., McMillan, A., Pingault, J.-B., et al. (2014). The high heritability of educational achievement reflects many genetically influenced traits, not just intelligence. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(42), 15273–15278. http://doi.org/10.1073/pnas.1408777111

Lee, J. J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., et al. (2018). Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature Genetics*, *50*(8), 1112–1121. http://doi.org/10.1038/s41588-018-0147-3

Locke, A. E., Kahali, B., Berndt, S. I., Justice, A. E., Pers, T. H., Day, F. R., et al. (2015). Genetic studies of body mass index yield new insights for obesity biology. *Nature*, *518*(7538), 197–U401. http://doi.org/10.1038/nature14177

Luciano, M., Hagenaars, S. P., Davies, G., Hill, W. D., Clarke, T.-K., Shirali, M., et al. (2018). Association analysis in over 329,000 individuals identifies 116 independent variants influencing neuroticism. *Nature Genetics*, *50*(1), 6–11. http://doi.org/10.1038/s41588-017-0013-8

Marques, S. C., Pais-Ribeiro, J. L., & Lopez, S. J. (2011). The Role of Positive Psychology Constructs in Predicting Mental Health and Academic Achievement in Children and Adolescents: A Two-Year Longitudinal Study. *Journal of Happiness Studies*, *12*(6), 1049–1062. http://doi.org/10.1007/s10902-010-9244-4

Martin, A. R., Daly, M. J., Robinson, E. B., Hyman, S. E., & Neale, B. M. (2018). Predicting polygenic risk of psychiatric disorders. *Biological Psychiatry*. http://doi.org/10.1016/j.biopsych.2018.12.015

McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A. R., Teumer, A., et al. (2016). A reference panel of 64,976 haplotypes for genotype imputation. *Nature Genetics*, *48*(10), 1279–1283. http://doi.org/10.1038/ng.3643

McInnes, G., Tanigawa, Y., DeBoever, C., Lavertu, A., Olivieri, J. E., Aguirre, M., & Rivas, M. A. (2018). Global Biobank Engine: enabling genotype-phenotype browsing for biobank summary statistics. *Bioinformatics*, *9*, 1612. http://doi.org/10.1093/bioinformatics/bty999

Mõttus, R., Realo, A., Vainik, U., Allik, J., & Esko, T. (2017). Educational Attainment and Personality Are Genetically Intertwined. *Psychological Science*, *28*(11), 1631–1639. http://doi.org/10.1177/0956797617719083

Mullins-Sweatt, S. N., Jamerson, J. E., Samuel, D. B., Olson, D. R., & Widiger, T. A. (2006). Psychometric properties of an abbreviated instrument of the five-factor model. *Assessment*, *13*(2), 119–137. http://doi.org/10.1177/1073191106286748

Pardiñas, A. F., Holmans, P., Pocklington, A. J., Escott-Price, V., Ripke, S., Carrera, N., et al. (2018). Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nature Genetics*, *50*(3), 381–389. http://doi.org/10.1038/s41588-018-0059-2

Paternoster, R., Brame, R., Mazerolle, P., & Piquero, A. (1998). Using the correct statistical test for the equality of regression coefficients. *Criminology*, *36*(4), 859–866. http://doi.org/10.1111/j.1745-9125.1998.tb01268.x

Pingault, J.-B., O'Reilly, P. F., Schoeler, T., Ploubidis, G. B., Rijsdijk, F., & Dudbridge, F. (2018). Using genetic data to strengthen causal inference in observational research. *Nature Reviews. Genetics*, *19*(9), 566–580. http://doi.org/10.1038/s41576-018-0020-3

Pingault, J.-B., Tremblay, R. E., Vitaro, F., Carbonneau, R., Genolini, C., Falissard, B., & Cote, S. M. (2011). Childhood Trajectories of Inattention and Hyperactivity and Prediction of Educational Attainment in Early Adulthood: A 16-Year Longitudinal Population-Based Study. *American Journal of Psychiatry*, *168*(11), 1164–1170. http://doi.org/10.1176/appi.ajp.2011.10121732

Plomin, R. (1994). Genetics and experience: The interplay between nature and nurture. Sage Publications, Inc.

Plomin, R. (2018). Blueprint. Penguin UK.

Plomin, R., & Stumm, von, S. (2018). The new genetics of intelligence. *Nature Reviews. Genetics*, *19*(3), 148–159. http://doi.org/10.1038/nrg.2017.104

Plomin, R., DeFries, J. C., & Loehlin, J. C. (1977). Genotype-environment interaction and correlation in the analysis of human behavior. *Psychological Bulletin*, *84*(2), 309–322. http://doi.org/10.1037/0033-2909.84.2.309

Polderman, T. J. C., Benyamin, B., de Leeuw, C. A., Sullivan, P. F., van Bochoven, A., Visscher, P. M., & Posthuma, D. (2015). Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nature Genetics*, *47*(7), 702–709. http://doi.org/10.1038/ng.3285

Raven, J., & Court, J. H. (1996). Manual for Raven's Progressive Matrices and Vocabulary Scales. Oxford: Oxford University Press.

Rietveld, C. A., Esko, T., Davies, G., Pers, T. H., Turley, P., Benyamin, B., et al. (2014). Common genetic variants associated with cognitive performance identified using the proxy-phenotype method. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(38), 13790–13794. http://doi.org/10.1073/pnas.1404623111

Rietz, Du, E., Cheung, C. H. M., McLoughlin, G., Brandeis, D., Banaschewski, T., Asherson, P., & Kuntsi, J. (2016). Self-report of ADHD shows limited agreement with objective markers of persistence and remittance. *Journal of Psychiatric Research*, *82*, 91–99. http://doi.org/10.1016/j.jpsychires.2016.07.020

Rijsdijk, F. V. (2002). Analytic approaches to twin data using structural equation models. *Briefings in Bioinformatics*, *3*(2), 119–133. http://doi.org/10.1093/bib/3.2.119

Savage, J. E., Jansen, P. R., Stringer, S., Watanabe, K., Bryois, J., de Leeuw, C. A., et al. (2018). Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nature Genetics*, *50*(7), 912–919. http://doi.org/10.1038/s41588-018-0152-6

Selzam, S., Krapohl, E., Stumm, von, S., O'Reilly, P. F., Rimfeld, K., Kovas, Y., et al. (2017). Predicting educational achievement from DNA. *Molecular Psychiatry*, *22*(2), 267–272. http://doi.org/10.1038/mp.2016.107

Selzam, S., McAdams, T. A., Coleman, J. R. I., Carnell, S., O'Reilly, P. F., Plomin, R., & Llewellyn, C. H. (2018). Evidence for gene-environment correlation in child feeding: Links between common genetic variation for BMI in children and parental feeding practices. *PLOS Genetics*, *14*(11), e1007757. http://doi.org/10.1371/journal.pgen.1007757

Shakeshaft, N. G., Trzaskowski, M., McMillan, A., Rimfeld, K., Krapohl, E., Haworth, C. M. A., et al. (2013). Strong Genetic Influence on a UK Nationwide Test of Educational Achievement at the End of Compulsory Education at Age 16. *Plos One*, *8*(12), e80341. http://doi.org/10.1371/journal.pone.0080341

Smith, G. D. (2007). Capitalizing on Mendelian randomization to assess the effects of treatments. *Journal of the Royal Society of Medicine*, *100*(9), 432–435. http://doi.org/10.1177/014107680710000923

Smith, G. D., & Ebrahim, S. (2003). "Mendelian randomization": can genetic epidemiology contribute to understanding environmental determinants of disease? *International Journal of Epidemiology*, *32*(1), 1–22. http://doi.org/10.1093/ije/dyg070

Smith, G. D., & Ebrahim, S. (2005). What can mendelian randomisation tell us about modifiable behavioural and environmental exposures? *Bmj*, *330*(7499), 1076–1079. http://doi.org/10.1136/bmj.330.7499.1076

Smith, G. D., & Hemani, G. (2014). Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Human Molecular Genetics*, *23*(R1), R89–R98. http://doi.org/10.1093/hmg/ddu328

Smith-Woolley, E., Selzam, S., & Plomin, R. (2019). Polygenic score for educational attainment captures DNA variants shared between personality traits and educational achievement. *Journal of Personality and Social Psychology: Personality Processes and Individual Differences*. http://doi.org/10.1037/pspp0000241

Spinath, B., Spinath, F. M., Harlaar, N., & Plomin, R. (2006). Predicting school achievement from general cognitive ability, self-perceived ability, and intrinsic value. *Intelligence*, *34*(4), 363–374. http://doi.org/10.1016/j.intell.2005.11.004

Trejo, S., & Domingue, B. W. (2019). Genetic Nature or Genetic Nurture? Quantifying Bias in Analyses Using Polygenic Scores: Appendix. *bioRxiv*, 524850. http://doi.org/10.1101/524850

Trzaskowski, M., Harlaar, N., Arden, R., Krapohl, E., Rimfeld, K., McMillan, A., et al. (2014). Genetic influence on family socioeconomic status and children's intelligence. *Intelligence*, *42*(100), 83–88. http://doi.org/10.1016/j.intell.2013.11.002

Vilhjalmsson, B. J., Yang, J., Finucane, H. K., Gusev, A., Lindstrom, S., Ripke, S., et al. (2015). Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. *American Journal of Human Genetics*, *97*(4), 576–592. http://doi.org/10.1016/j.ajhg.2015.09.001

Ware, J. E., & Sherbourne, C. D. (1992). The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Medical Care*, *30*(6), 473–483.

Wechsler, D. (1992). Wechsler Intelligence Scale for Children - Third Edition UK (WISC-III-UK) Manual. London: The Psychological Corporation.

Wood, A. R., Esko, T., Yang, J., Vedantam, S., Pers, T. H., Gustafsson, S., et al. (2014). Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature Genetics*, *46*(11), 1173–1186. http://doi.org/10.1038/ng.3097

Wray, N. R., Lee, S. H., Mehta, D., Vinkhuyzen, A. A. E., Dudbridge, F., & Middeldorp, C. M. (2014). Research review: Polygenic methods and their application to psychiatric traits. *Journal of Child Psychology and Psychiatry*, *55*(10), 1068–1087. http://doi.org/10.1111/jcpp.12295

Wray, N. R., Yang, J., Ben J Hayes, Price, A. L., Goddard, M. E., & Visscher, P. M. (2013). Pitfalls of predicting complex traits from SNPs. *Nature Reviews. Genetics*, *14*(7), 507–515. http://doi.org/10.1038/nrg3457

Yang, J., Lee, S. H., Goddard, M. E., & Visscher, P. M. (2011). GCTA: A Tool for Genome-wide Complex Trait Analysis. *The American Journal of Human Genetics*, *88*(1), 76–82. http://doi.org/10.1016/j.ajhg.2010.11.011

Yengo, L., Sidorenko, J., Kemper, K. E., Zheng, Z., Wood, A. R., Weedon, M. N., et al. (2018). Meta-analysis of genome-wide association studies for height and body mass index in similar to 700 000 individuals of European ancestry. *Human Molecular Genetics*, *27*(20), 3641–3649. http://doi.org/10.1093/hmg/ddy271

Young, A. I., Frigge, M. L., Gudbjartsson, D. F., Thorleifsson, G., Bjornsdottir, G., Sulem, P., et al. (2018). Relatedness disequilibrium regression estimates heritability without environmental bias. *Nature Genetics*, *50*(9), 1304–1310. http://doi.org/10.1038/s41588-018-0178-9

Zuffianò, A., Alessandri, G., Gerbino, M., Luengo Kanacri, B. P., Di Giunta, L., Milioni, M., & Caprara, G. V. (2013). Academic achievement: The unique contribution of self-efficacy beliefs in self-regulated learning beyond intelligence, personality traits, and self-esteem. *Learning and Individual Differences*, *23*, 158–162. http://doi.org/10.1016/j.lindif.2012.07.010

Chapter 7 – General Summary and Discussion

Polygenic score analysis has become a popular tool for the investigation of complex human traits. It was the aim of this thesis to investigate developmental and multivariate research questions, as well as genotype-environment interplay through the use of a broad range of polygenic scores analysis approaches and phenotypes.

This chapter summarises the key findings from the preceding empirical chapters, discusses general limitations of the polygenic score approach, elaborates on the broader implications of these findings, and provides possible future directions relating to within-family polygenic score analysis.

Summary of findings

With increasing statistical power to detect genetic effects between a single nucleotide polymorphism (SNP) and trait outcomes due to growing GWA study sample sizes, genome-wide polygenic score (GPS) analysis has become more powerful. In Chapter 2, a GPS based on the largest GWA study for years of education at the time (Okbay et al., 2016), was shown to predict educational achievement across development, explaining up to 9.1% of the trait variance at age 16. Quantile analyses of the polygenic predictor on educational achievement at age 16 indicated a whole school grade difference between the extreme groups. Furthermore, the years of education GPS predicted 3.6% of the variance in intelligence, and 7.3% of the variance in family socioeconomic status (SES). A statistical model applied to test the interaction between the GPS and SES on (i) educational achievement at age 16 and (ii) intelligence, did not suggest significant interaction effects.

Educational achievement is not only a product of an individual's intelligence, but also relates to personality and motivation. Chapter 3 set out to explore the relationship between the years of education GPS and six personality and motivation domains, and was found to predict between 0.6% and 3% of the trait variances. Even after controlling for intelligence, most of these associations remained significant. Structural equation modelling was applied to test the extent to which the GPS accounted for the phenotypic covariance between the six domains and educational achievement at age 16, which was found to range between 8% and 16%. Results relating to the education GPS were contrasted to two personality GPS for neuroticism and wellbeing, which were found to be less predictive, explaining between 0.3% and 0.7% of the trait variances. The results in Chapter 2 and 3 highlight the strength of the

years of education GPS – also relative to other GPS – and illustrate the substantial pleiotropy across education and related traits.

Extensive pleiotropy has also been observed across different forms of psychiatric disorders, which may be due to their overlapping clinical symptoms and diagnostic criteria. Therefore, Chapter 4 investigated a shared genetic psychopathology (*p*) factor between psychiatric disorders, by applying principal component analysis (PCA) on a correlation matrix of GPS for psychiatric disorders, and on three genetic correlation matrices derived from three genetic methods: linkage disequilibrium score regression (LDSC), genome-wide complex trait analysis (GCTA-REML), and family analysis. Results from PCA showed that the first principal component explained 22% of the GPS data, 43% of the LDSC data, 35% of the GCTA data, and 57% of the family analysis data. Furthermore, all genetic loadings on the first principal component were positive across all four methods, indicating that the genetic effects for each psychiatric disorder are positively correlated with the latent genetic dimension. Overall, findings from Chapter 4 provided strong support for a genetic *p* factor that indexes a general genetic liability to the development of psychiatric disorders.

Chapter 5 and 6 were concerned more specifically with identifying genotype-environment correlation (rGE) through GPS analysis. In Chapter 5, evocative rGE in the home environment was tested for the relationship between the offspring's GPS for BMI and their parental feeding practices. In unrelated individuals, the GPS was positively associated with parental restriction over food intake, and negatively associated with parental pressure to eat more food. The magnitudes of effect sizes were almost identical within families when performing fixed-effect analysis based on dizygotic (DZ) twin pair data, indicating that the observed effect is offspring-driven. Multivariate twin analysis was also employed, and the directions of estimated genetic correlations between child BMI and parental feeding practices were consistent with those found in GPS analyses.

Another possible application of GPS analysis is the estimation of the passive rGE effect in unrelated (i.e. between-family) samples. Parents generate family environments consistent with their own genotypes, which in turn facilitate the development of the offspring trait, thus inducing a correlation between offspring genotype and family environment. Therefore, GPS prediction among unrelated individuals may include contributions from both direct genetic effects and also indirect effects due to prGE. In Chapter 6, we applied a random intercept

mixed-effects model to DZ twin pair data to simultaneously estimate within- and between-family GPS predictions for eight GPS predicting each of the eight traits included in the study. Target- and cross-trait analyses indicated that polygenic score prediction estimates for cognitive traits (intelligence and educational achievement) were on average 60% greater between families than within families, smaller, non-significant differences were found for non-cognitive traits. The within- and between-family difference for cognitive traits disappeared after controlling for SES, suggesting that SES is a major source of between-family prediction through rGE mechanisms. Overall, findings from Chapter 5 and 6 illustrate the different ways in which GPS analysis can be utilised to investigate the genetic mediation of environmental influences on trait development.

General limitations

Limitations specific to each study were discussed within each chapter. The following limitations presented are general to the polygenic score approach.

Predictive power of polygenic scores

The ability to predict trait variance in samples using the GPS approach depends on, and is limited by, several factors. The maximum possible prediction value is determined by the extent to which trait variability in a population is explained by genetic variability (i.e. heritability; h^2). Although $h^2_{pedigree}$ is considered as the upper threshold for additive genetic effects, h^2_{SNP} represents the ceiling for GPS prediction as it is estimated using directly measured or imputed SNPs. Currently, h^2_{SNP} is generally half that of $h^2_{pedigree}$ and GPS predictions are generally substantially less than h^2_{SNP} (Wray, Kemper, Hayes, Goddard, & Visscher, 2019). This difference can be due to (i) the genetic difference between target sample and the discovery sample; (ii) the genetic architecture of the trait, (iii) the target sample size, and (iv) the power of the GWA study to accurately estimate individual SNP effects, which relates to the measurement errors which are then summed up and included in the GPS (Dudbridge, 2013; 2016; Palla & Dudbridge, 2015; Wray et al., 2019; Wray, Goddard, & Visscher, 2007).

With increasing GWA study sample sizes, the predictive accuracy of GPSs has improved steadily. Therefore, one strategy to improve polygenic trait prediction is to continuously increase the GWA study sample sizes as more genotype data become available.

A further challenge will be to reduce the missing heritability gap between h^2_{SNP} and $h^2_{pedigree}$ to push up the ceiling for GPS predictions. A recent study that applied the GCTA-GREML approach on whole genome sequencing (WGS) data found that the inclusion of rare genetic variation (0.0001 < MAF < 0.001) in low LD accounted for the previously missing proportion of heritability in height and BMI (Wainschtein et al., 2019). This could mean that GPS predictions may not improve considerably until GWA studies are performed on large WGS data, and information about rare variation is incorporated in the polygenic predictors. Although it has been shown that WGS data can recover much of the missing h^2_{SNP} for height and BMI, it is unclear whether this will apply similarly for behavioural traits, such as educational achievement and psychopathology.

Another potentially important aspect for increasing h^2_{SNP} may be to direct more focus on the validity of phenotype measurements. For example, h^2_{SNP} is generally low for psychiatric disorders (~10%) in samples that are well-powered to detect genetic effects (Cheesman et al., 2017; Ge, Chen, Neale, Sabuncu, & Smoller, 2017). As shown in Chapter 4 – and in more recently emerging research – evidence pointing towards a general genetic liability for psychiatric disorder is amassing (Allegrini et al., 2019; Brikell et al., 2018; Grotzinger et al., 2019; Riglin et al., 2018), which increasingly brings the current classification systems into question. As a result, h^2_{SNP}, and thus GPS predictions, might not increase until it is possible to capture general and specific risk factors for psychiatric diseases that are empirically derived (Conway et al., 2019).

Potential overestimation of polygenic score predictions

Two potential causes of upward bias in GPS predictions are data overfitting and the presence of sample overlap or cryptic relatedness between discovery and target samples. Overfitting may occur when performing parameter optimisation for the calculation of a GPS in the target sample, and then applying the most predictive GPS in the same sample (Choi, Mak, & O'Reilly, 2018). Out-of-sample prediction is a useful approach to guard against overfitting, which requires a separate training sample for parameter optimisation and a test sample for GPS analysis. If an independent sample is not available, k-fold cross-validation can be

applied (Kohavi, 1995). Here, the sample is repeatedly subdivided into training and validation data sets with different partitions, resulting in more stable prediction estimates, as implemented in Chapter 3.

Another main source of upward bias in polygenic prediction occurs in the presence of cryptic relatedness between the target and discovery GWA study sample (Choi et al., 2018; Makowsky et al., 2011; Wray et al., 2013). To test for cryptic relatedness between discovery and target samples in order to remove individuals with close relatedness, individual-level genotypic data are required. However, these data are most commonly not available for the discovery sample, and currently no statistical approaches exist to estimate, and adjust for, cryptic relatedness based on GWA study summary statistics and individual-level genotype target data. With ever-growing GWA study sample sizes, this could become an increasingly pressing issue, as it will become gradually more difficult to ensure that discovery and target samples are independent.

Generalisability of polygenic prediction to other populations

A current major challenge to GPS prediction is that it is inaccurate when applied to individuals from populations that are ancestrally different to the discovery sample. At present, around 78% of individuals used for GWA analysis are of white European ancestry (Morales et al., 2018). A further 11% of individuals are of Asian ancestry, and only 2.4% and 1.3% are of African or Hispanic/Latin American ancestry, respectively (Morales et al., 2018). In addition to the highly skewed availability of samples with European ancestry, ascertainment bias of SNP selection for microarray platforms are also largely based on European reference samples, potentially leading to risk misestimation (Kim, Patel, Teng, Berens, & Lachance, 2018; La Vega & Bustamante, 2018). The lack of diversity is highly problematic, as the accuracy of GPS prediction decreases as a function of population differences relating to (i) allele frequencies, and (ii) linkage disequilibrium (Martin et al., 2017; 2019; Scutari, Mackay, & Balding, 2016). For example, in contrast to white European samples: the predictive accuracy of the years of education GPS was 85% lower in African Americans (Lee et al., 2018); the predictive accuracy of a schizophrenia GPS was attenuated by 90% in African samples (Vassos et al., 2017); and for GPSs relating to anthropometric and blood-related traits, predictions were on average ~45% lower in South Asians, ~45% lower in Hispanic/Latino Americans, ~50% lower in East Asians, and ~75% lower in Africans (Martin

et al., 2019). These current systematic biases can lead to substantial disparities if polygenic prediction is applied to healthcare settings (La Vega & Bustamante, 2018; Martin et al., 2019; Popejoy & Fullerton, 2016).

The key to reducing and eliminating this bias towards white European samples is to prioritise the ascertainment of large, ancestrally diverse populations. This will make it possible to perform ancestry specific GWA studies as the foundation of more predictive GPS for currently understudied populations. Early promising results are starting to emerge, showing that predictive accuracy of GPSs are consistently higher in independent samples if GWA study populations are of the same ancestry (Martin et al., 2019). Efforts to increase ancestrally diverse samples are underway, for example the Trans-Omics for Precision Medicine (TOPMED) which holds WGS data for ~ 45,000 African Americans and for ~24,000 Hispanics or Latin Americans (https://www.nhlbiwgs.org), or the East London Genes & Health project, which currently includes genotype data for ~35,000 participants of South Asian ancestry with the goal to reach 100,000 (Finer et al., 2019).

The availability of ancestrally diverse samples will also aid the identification of causal variants. Variants that are rare in European samples may be common in other populations, and it has been demonstrated that GWA analysis of non-European samples can be considerably more powered to detect effects of causal disease variants that are rare in European samples (Estrada et al., 2014; The SIGMA Type 2 Diabetes Consortium, 2013). Ultimately, WGS GWA studies could provide most informative data as it also circumvents microarray SNP selection bias (La Vega & Bustamante, 2018).

The generalisability of GPS prediction does not only relate to genetic differences across populations, but also environmental differences. The ceiling for GPS prediction is h^2_{SNP}; however, h^2 is a population estimate that is directly influenced by environmental variance within that population (Mostafavi, Harpak, Conley, Pritchard, & Przeworski, 2019). For example, both h^2_{SNP} and GPS predictions of educational attainment were twice as great in Estonia for individuals brought up during the post-Soviet era in contrast to individuals brought up during the Soviet era due to diminished equality of opportunity during the communist regime (Rimfeld et al., 2018).

Overall, findings presented as part of this thesis may only be generalisable to the white European British population at the specific ages tested, and may not replicate directly in other populations with different education, social, or political settings.

Implications and future directions

In the following section, general implications and possible future directions will be presented in relation to this thesis.

Polygenic scores for individual trait prediction

The low cost of genotyping means that genotypic data are more widely available for research, but also to the individual through direct-to-consumer testing. One of the major promises of GPS prediction is that it can be used to identify individual risk and resilience. However, it is important to point out the constraints of individual GPS trait prediction, which is limited by h^2_{SNP}, the generally low predictive power of the polygenic predictor based on the various factors discussed previously, as well as unsystematic environmental effects that influence trait outcomes. These factors make achieving accurate individual-specific GPS prediction a challenging endeavor.

A more promising approach is to stratify individuals from the study sample into groups based on their GPS values, and to draw conclusions about the groups that are at the extremes of the distribution based on their average trait values or probabilities. In Chapter 2 it was shown that there was a whole school grade difference between individuals that fell within the lowest and the highest ~15% of the years of education GPS distribution. In Chapter 6, it was demonstrated that this approach can also be applied to perform within-family comparisons of how GPS differences relate to trait differences within DZ twin pairs. For example, DZ twin pairs that fell into the highest 10% of the within-pair height GPS difference distribution showed an average 8.7 cm height difference in comparison to the twin pairs that fell into the lowest 10% of the GPS difference distribution.

Stratification based on GPS values has also become a popular approach in polygenic score health research (Chatterjee, Shi, & García-Closas, 2016; Chatterjee et al., 2013; Lewis & Vassos, 2017; Maher, 2015; Torkamani, Wineinger, & Topol, 2018). The power of this

approach has been demonstrated in a recent study that estimated the proportion of the UK Biobank sample that was at a three-fold risk to develop coronary artery disease (CAD) (8%), atrial fibrillation (6.1%), type 2 diabetes (3.5%), inflammatory bowel disease (3.2%), and breast cancer (1.5%) (Khera et al., 2018). The value of polygenic risk stratification becomes particularly apparent if compared to risk prediction based on single genetic variation. For example, the above study illustrated that the CAD GPS identified 20 times more individuals that are at the same or greater risk of developing CAD than single hypercholesterolemia mutations that relate to a three-fold increased risk (Khera et al., 2018). Taken together with wider information about risk factors relating to the environment, age and sex, polygenic scores have the potential to turn precision healthcare into reality (Chatterjee et al., 2016; Torkamani et al., 2018).

With polygenic scores rapidly gaining predictive power, it will be imperative to increase the awareness and understanding about complex trait genetics amongst the public and healthcare professionals if GPS were to be incorporated in healthcare (McBride, Koehly, Sanderson, & Kaphingst, 2010; McLaughlin et al., 2014; Sabatello, Chen, Sanderson, Chung, & Appelbaum, 2018; Vassy et al., 2014). It is crucial that information about genetic risk can be delivered in an effective and accessible way (Kong et al., 2014; McLaughlin et al., 2014), as this may determine the level of behavioural adjustment in individuals in response to their genetic risk (Sanderson, Waller, Humphries, & Wardle, 2011).

Combining within-family polygenic score variation with family risk information

Family disease history is currently one of the most widely used risk indicator in clinical settings (Kong et al., 2014), and is a rough measure of polygenic and environmental risk (Dudbridge, Pashayan, & Yang, 2017). Few studies set out to test the clinical utility of GPS beyond known family risk in samples of unrelated individuals. For a schizophrenia GPS, a significant interaction with family history was found, showing that family history only related to substantially higher schizophrenia odds ratios in the top three GPS deciles (Agerbo et al., 2015). In contrast, the prediction of heart disease using a CAD GPS was largely independent of family risk, highlighting the potential utility of using GPS information in addition to family history (Tada et al., 2016). However, the polygenic score only included 50 SNPs, and family history was self-reported, potentially lacking accuracy.

Further research is needed to determine the utility of within-family GPS prediction when used in combination with known family risk. Although family risk applies equally to all siblings within a family, their individual-level genetic risk may differ considerably as their identity by descent is on average 0.50, with an expected range at -/+ 2 SDs of 0.40 – 0.60 (Visscher et al., 2006). As shown in Chapter 6, within-family GPS variation can predict trait differences between DZ co-twins for various outcomes. Although no clinical data was available for these analyses, the mixed-effects model can be applied to other datasets such as the UK Biobank, which includes clinical data for over 22,000 sibling pairs (Bycroft et al., 2018), as well as data on family history for a range of diseases. To evaluate an improvement in disease prediction, a baseline model containing only family history status as a fixed effect and a family-level random effect could be compared to the model applied in Chapter 6 with the addition of the family history fixed effect, and without the between-family fixed effect as the interest is the within-family prediction beyond family risk. If the addition of the within-family GPS fixed effect produces a significantly higher coefficient of determination (R^2) than the baseline model, within-family GPS prediction may be a useful addition to family history information to inform individual-level disease risk. Depending on the data availability, other fixed effects relating to extended family history information can be included, for example parental age of onset of disease.

Using multivariate approaches to maximise within-family variance explained

Multivariate approaches to polygenic prediction are continuously developed to improve predictive accuracy by leveraging the pervasive pleiotropy across complex traits. For example, regularized regression in form of elastic net has been shown to significantly improve prediction of cognitive traits and BMI by incorporating over 80 GPS into the model (Krapohl et al., 2018). Another approach is to incorporate genetic information at the summary statistic level before calculating the GPS, by jointly analysing single SNP effects across the genome for multiple, related traits using structural equation modelling (Grotzinger et al., 2019). This method increases statistical power for genetic discovery, and GPS derived from the resulting summary statistics have been shown to outperform single-trait GPS (Grotzinger et al., 2019).

These approaches have so far only been applied to data from unrelated samples. As shown in Chapter 6, within-family GPS effect sizes can be considerably smaller than between-

family effects for the prediction of cognitive traits due to effects related to passive genotype-environment correlation, population stratification and/or assortative mating. Therefore, it will be crucial to employ multivariate analyses to improve within-family GPS prediction. A systematic comparison between various multivariate approaches would be insightful as to which approach most effectively maximises within-family trait prediction.

Conclusion

In conclusion, the work presented as part of this thesis addressed developmental and multivariate research questions, as well as genotype-environment interplay through the use of a broad range of polygenic scores analysis approaches and phenotypes. The findings of this thesis illustrate the potential of polygenic score analysis across a wide range of complex traits. In combination with a variety of patient-specific information, polygenic scores may be applied in clinical settings in the near future. However, until then, immense challenges will need to be overcome, relating to the predictive power of polygenic scores, the lack of diversity, and the public understanding of genomic information. Within-family polygenic score analysis represents a promising area for future research.

References

Agerbo, E., Sullivan, P. F., Vilhjalmsson, B. J., Pedersen, C. B., Mors, O., Børglum, A. D., et al. (2015). Polygenic Risk Score, Parental Socioeconomic Status, Family History of Psychiatric Disorders, and the Risk for Schizophrenia A Danish Population-Based Study and Meta-analysis. *Jama Psychiatry*, *72*(7), 635–641. http://doi.org/10.1001/jamapsychiatry.2015.0346

Allegrini, A. G., Cheesman, R., Rimfeld, K., Selzam, S., Pingault, J.-B., Eley, T. C., & Plomin, R. (2019). The pfactor: Genetic analyses support a general dimension of psychopathology in childhood and adolescence. *bioRxiv*, 591354. http://doi.org/10.1101/591354

Brikell, I., Larsson, H., Lu, Y., Pettersson, E., Chen, Q., Kuja-Halkola, R., et al. (2018). The contribution of common genetic risk variants for ADHD to a general factor of childhood psychopathology. *Molecular Psychiatry*, *1*, 1–13. http://doi.org/10.1038/s41380-018-0109-2

Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L. T., Sharp, K., et al. (2018). The UK Biobank resource with deep phenotyping and genomic data. *Nature*, *562*(7726), 203–209. http://doi.org/10.1038/s41586-018-0579-z

Chatterjee, N., Shi, J., & García-Closas, M. (2016). Developing and evaluating polygenic risk prediction models for stratified disease prevention. *Nature Reviews. Genetics*, *17*(7), 392–406. http://doi.org/10.1038/nrg.2016.27

Chatterjee, N., Wheeler, B., Sampson, J., Hartge, P., Chanock, S. J., & Park, J.-H. (2013). Projecting the performance of risk prediction based on polygenic analyses of genome-wide association studies. *Nature Genetics*, *45*(4), 400–405. http://doi.org/10.1038/ng.2579

Cheesman, R., Selzam, S., Ronald, A., Dale, P. S., McAdams, T. A., Eley, T. C., & Plomin, R. (2017). Childhood behaviour problems show the greatest gap between DNA-based and twin heritability. *Translational Psychiatry*, *7*(12), 1284. http://doi.org/10.1038/s41398-017-0046-x

Choi, S. W., Mak, T. S. H., & O'Reilly, P. (2018). A guide to performing Polygenic Risk Score analyses. *bioRxiv*, 416545. http://doi.org/10.1101/416545

Conway, C. C., Forbes, M. K., Forbush, K. T., Fried, E. I., Hallquist, M. N., Kotov, R., et al. (2019). A Hierarchical Taxonomy of Psychopathology Can Transform Mental Health Research. *Perspectives on Psychological Science : a Journal of the Association for Psychological Science*, *80*, 1745691618881069. http://doi.org/10.1177/1745691618810696

Dudbridge, F. (2013). Power and Predictive Accuracy of Polygenic Risk Scores. *PLOS Genetics*, *9*(3). http://doi.org/10.1371/journal.pgen.1003348

Dudbridge, F. (2016). Polygenic Epidemiology. *Genetic Epidemiology*, *40*(4), 268–272. http://doi.org/10.1002/gepi.21966

Dudbridge, F., Pashayan, N., & Yang, J. (2017). Predictive accuracy of combined genetic and environmental risk scores. *Genetic Epidemiology*, *42*(1), 4–19. http://doi.org/10.1002/gepi.22092

Estrada, K., Aukrust, I., Bjørkhaug, L., Burtt, N. P., Mercader, J. M., García-Ortiz, H., et al. (2014). Association of a Low-Frequency Variant in HNF1A With Type 2 Diabetes in a Latino Population. *Jama*, *311*(22), 2305–2314. http://doi.org/10.1001/jama.2014.6511

Finer, S., Martin, H., Hunt, K. A., MacLaughlin, B., Ashcroft, R., Khan, A., et al. (2019). Cohort Profile: East London Genes & Health (ELGH), a community based population genomics and health study of British-Bangladeshi and British-Pakistani people. *bioRxiv*, 426163. http://doi.org/10.1101/426163

Ge, T., Chen, C.-Y., Neale, B. M., Sabuncu, M. R., & Smoller, J. W. (2017). Phenome-wide heritability analysis of the UK Biobank. *PLOS Genetics*, *13*(4), e1006711. http://doi.org/10.1371/journal.pgen.1006711

Grotzinger, A. D., Rhemtulla, M., de Vlaming, R., Ritchie, S. J., Mallard, T. T., Hill, W. D., et al. (2019). Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. *Nature Human Behaviour*, *45*, 984. http://doi.org/10.1038/s41562-019-0566-x

Khera, A. V., Chaffin, M., Aragam, K. G., Haas, M. E., Roselli, C., Choi, S. H., et al. (2018). Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nature Genetics*, *50*(9), 1219–1224. http://doi.org/10.1038/s41588-018-0183-z

Kim, M. S., Patel, K. P., Teng, A. K., Berens, A. J., & Lachance, J. (2018). Genetic disease risks can be misestimated across global populations. *Genome Biology*, *19*(1), 179. http://doi.org/10.1186/s13059-018-1561-7

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, *2*, 1137–1145.

Kong, S. W., Lee, I.-H., Leshchiner, I., Krier, J., Kraft, P., Rehm, H. L., et al. (2014). Summarizing polygenic risks for complex diseases in a clinical whole-genome report. *Genetics in Medicine : Official Journal of the American College of Medical Genetics*, *17*(7), 536–544. http://doi.org/10.1038/gim.2014.143

Krapohl, E., Patel, H., Newhouse, S., Curtis, C. J., Stumm, von, S., Dale, P. S., et al. (2018). Multi-polygenic score approach to trait prediction. *Molecular Psychiatry*, *23*(5), 1368–1374. http://doi.org/10.1038/mp.2017.163

La Vega, De, F. M., & Bustamante, C. D. (2018). Polygenic risk scores: a biased prediction? *Genome Medicine*, *10*(1), 100. http://doi.org/10.1186/s13073-018-0610-x

Lee, J. J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., et al. (2018). Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature Genetics*, *50*(8), 1112–1121. http://doi.org/10.1038/s41588-018-0147-3

Lewis, C. M., & Vassos, E. (2017). Prospects for using risk scores in polygenic medicine. *Genome Medicine*, *9*(1), 1466. http://doi.org/10.1186/s13073-017-0489-y

Maher, B. S. (2015). Polygenic Scores in Epidemiology: Risk Prediction, Etiology, and Clinical Utility. *Current Epidemiology Reports*, *2*(4), 239–244. http://doi.org/10.1007/s40471-015-0055-3

Makowsky, R., Pajewski, N. M., Klimentidis, Y. C., Vazquez, A. I., Duarte, C. W., Allison, D. B., & de los Campos, G. (2011). Beyond Missing Heritability: Prediction of Complex Traits. *PLOS Genetics*, *7*(4), e1002051. http://doi.org/10.1371/journal.pgen.1002051

Martin, A. R., Gignoux, C. R., Walters, R. K., Wojcik, G. L., Neale, B. M., Gravel, S., et al. (2017). Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. *The American Journal of Human Genetics*, *100*(4), 635–649. http://doi.org/10.1016/j.ajhg.2017.03.004

Martin, A. R., Kanai, M., Kamatani, Y., Okada, Y., Neale, B. M., & Daly, M. J. (2019). Clinical use of current polygenic risk scores may exacerbate health disparities. *Nature Genetics*, *51*(4), 584–591. http://doi.org/10.1038/s41588-019-0379-x

McBride, C. M., Koehly, L. M., Sanderson, S. C., & Kaphingst, K. A. (2010). The Behavioral Response to Personalized Genetic Information: Will Genetic Risk Profiles Motivate Individuals and Families to Choose More Healthful Behaviors? *Annual Review of Public Health*, *31*(1), 89–103. http://doi.org/10.1146/annurev.publhealth.012809.103532

McLaughlin, H. M., Ceyhan-Birsoy, O., Christensen, K. D., Kohane, I. S., Krier, J., Lane, W. J., et al. (2014). A systematic approach to the reporting of medically relevant findings from whole genome sequencing. *BMC Medical Genetics*, *15*(1), 2418. http://doi.org/10.1186/s12881-014-0134-1

Morales, J., Welter, D., Bowler, E. H., Cerezo, M., Harris, L. W., McMahon, A. C., et al. (2018). A standardized framework for representation of ancestry data in genomics studies, with

application to the NHGRI-EBI GWAS Catalog. *Genome Biology*, *19*(1), 489. http://doi.org/10.1186/s13059-018-1396-2

Mostafavi, H., Harpak, A., Conley, D., Pritchard, J. K., & Przeworski, M. (2019). Variable prediction accuracy of polygenic scores within an ancestry group. *bioRxiv*, *51*(4), 636. http://doi.org/10.1101/629949

Okbay, A., Beauchamp, J. P., Fontana, M. A., Lee, J. J., Pers, T. H., Rietveld, C. A., et al. (2016). Genome-wide association study identifies 74 loci associated with educational attainment. *Nature*, *533*(7604), 539–542. http://doi.org/10.1038/nature17671

Palla, L., & Dudbridge, F. (2015). A Fast Method that Uses Polygenic Scores to Estimate the Variance Explained by Genome-wide Marker Panels and the Proportion of Variants Affecting a Trait. *American Journal of Human Genetics*, *97*(2), 250–259. http://doi.org/10.1016/j.ajhg.2015.06.005

Popejoy, A. B., & Fullerton, S. M. (2016). Genomics is failing on diversity. *Nature*, *538*(7624), 161–164. http://doi.org/10.1038/538161a

Riglin, L., Thapar, A. K., Leppert, B., Martin, J., Richards, A., Anney, R., et al. (2018). Using genetics to examine a general liability to childhood psychopathology:. *bioRxiv*, 409540. http://doi.org/10.1101/409540

Rimfeld, K., Krapohl, E., Trzaskowski, M., Coleman, J. R. I., Selzam, S., Dale, P. S., et al. (2018). Genetic influence on social outcomes during and after the Soviet era in Estonia. *Nature Human Behaviour*, *2*(4), 269–275. http://doi.org/10.1038/s41562-018-0332-5

Sabatello, M., Chen, Y., Sanderson, S. C., Chung, W. K., & Appelbaum, P. S. (2018). Increasing genomic literacy among adolescents. *Genetics in Medicine : Official Journal of the American College of Medical Genetics*, *21*(4), 994–1000. http://doi.org/10.1038/s41436-018-0275-2

Sanderson, S. C., Waller, J., Humphries, S. E., & Wardle, J. (2011). Public Awareness of Genetic Influence on Chronic Disease Risk: Are Genetic and Lifestyle Causal Beliefs Compatible? *Public Health Genomics*, *14*(4-5), 290–297. http://doi.org/10.1159/000294280

Scutari, M., Mackay, I., & Balding, D. (2016). Using Genetic Distance to Infer the Accuracy of Genomic Prediction. *PLOS Genetics*, *12*(9), e1006288. http://doi.org/10.1371/journal.pgen.1006288

Tada, H., Melander, O., Louie, J. Z., Catanese, J. J., Rowland, C. M., Devlin, J. J., et al. (2016). Risk prediction by genetic risk scores for coronary heart disease is independent of self-reported family history. *European Heart Journal*, *37*(6), 561–567. http://doi.org/10.1093/eurheartj/ehv462

The SIGMA Type 2 Diabetes Consortium. (2013). Sequence variants in SLC16A11 are a common risk factor for type 2 diabetes in Mexico. *Nature*, *506*(7486), 97–101. http://doi.org/10.1038/nature12828

Torkamani, A., Wineinger, N. E., & Topol, E. J. (2018). The personal and clinical utility of polygenic risk scores. *Nature Reviews. Genetics*, *19*(9), 581–590. http://doi.org/10.1038/s41576-018-0018-x

Vassos, E., Di Forti, M., Coleman, J., Iyegbe, C., Prata, D., Euesden, J., et al. (2017). An Examination of Polygenic Score Risk Prediction in Individuals With First-Episode Psychosis. *Biological Psychiatry*, *81*(6), 470–477. http://doi.org/10.1016/j.biopsych.2016.06.028

Vassy, J. L., Lautenbach, D. M., McLaughlin, H. M., Kong, S. W., Christensen, K. D., Krier, J., et al. (2014). The MedSeq Project: a randomized trial of integrating whole genome sequencing into clinical medicine. *Trials*, *15*(1), 85. http://doi.org/10.1186/1745-6215-15-85

Visscher, P. M., Medland, S. E., Ferreira, M. A. R., Morley, K. I., Zhu, G., Cornes, B. K., et al. (2006). Assumption-Free Estimation of Heritability from Genome-Wide Identity-by-Descent Sharing between Full Siblings. *PLOS Genetics*, *2*(3), e41. http://doi.org/10.1371/journal.pgen.0020041

Wainschtein, P., Jain, D. P., Yengo, L., Zheng, Z., TOPMed Anthropometry Working Group, Trans-Omics for Precision Medicine Consortium, Cupples, L. A., et al. (2019). Recovery of trait heritability from whole genome sequence data. *bioRxiv*, 588020. http://doi.org/10.1101/588020

Wray, N. R., Goddard, M. E., & Visscher, P. M. (2007). Prediction of individual genetic risk to disease from genome-wide association studies. *Genome Research*, *17*(10), 1520–1528. http://doi.org/10.1101/gr.6665407

Wray, N. R., Kemper, K. E., Hayes, B. J., Goddard, M. E., & Visscher, P. M. (2019). Complex Trait Prediction from Genome Data: Contrasting EBV in Livestock to PRS in Humans. *Genetics*, *211*(4), 1131–1141. http://doi.org/10.1534/genetics.119.301859

Wray, N. R., Yang, J., Ben J Hayes, Price, A. L., Goddard, M. E., & Visscher, P. M. (2013). Pitfalls of predicting complex traits from SNPs. *Nature Reviews. Genetics*, *14*(7), 507–515. http://doi.org/10.1038/nrg3457

Appendices

The supplementary materials for several chapters, as referenced in the text, are attached as appendices:

Appendix 1 – Supplementary Material for Chapter 2

Appendix 2 – Supplementary Material for Chapter 3

Appendix 3 – Supplementary Material for Chapter 4

Appendix 4 – Supplementary Material for Chapter 5

Appendix 5 – Supplementary Material for Chapter 6

Appendix 1

Predicting Educational Achievement from DNA

SUPPLEMENTARY TABLES

SUPPLEMENTARY FIGURES

SUPPLEMENTARY METHODS

Supplementary Tables

Table S1. Representativeness of genotyped TEDS sample

	N	Female	A-levels: parental	Mother employed	Father employed
UK census 2001	N/A	50%	32%	49%	89%
TEDS 1st contact	5,825	53%	33%	46%	93%

Note. We used the 2001 UK census data (ONS, 2001; https://www.ons.gov.uk/census/2001censusandearlier/aboutcensus2001) for families with children rather than 2011 UK census data for TEDS twins as they were born between 1994-96.

Table S2. Descriptive Statistics

	N	Mean	SE	Mean F	SE F	Mean M	SE M	skew	min	max	R^2	$R^{2\prime}$
EA 7*	4047	0.05	0.02	0.08	0.02	0.01	0.02	-0.49	-3.96	2.88	<.01	0.02
EA 12*	2950	0.03	0.02	0.07	0.02	-0.02	0.03	0.4	-3.52	4.85	<.01	0.09
EA 16	4301	8.92	0.02	9.01	0.03	8.84	0.03	-0.49	4	11	<.01	<.01
SES*	4958	0.12	0.01	0.14	0.02	0.18	0.02	0.07	-2.6	2.58	N/A	N/A
g comp*	2228	0.11	0.02	0.15	0.03	0.09	0.02	-0.07	-2.71	2.55	nil	nil
g 7*	3559	0.04	0.02	0.05	0.02	0.03	0.03	-0.1	-4	5	nil	<.01
g 12*	3349	0.04	0.02	-0.04	0.02	0.13	0.03	-0.31	-3	3	<.01	0.05
g 16*	1743	0	0.02	-0.03	0.03	0.04	0.04	0.28	-3	3	<.01	<.01

Note. *N* = number of participants; R^2 = variance explained by gender differences; $R^{2\prime}$ = variance explained by age; EA = educational achievement; SES = socioeconomic status; *g* comp = general cognitive ability composite based on *g* measures at age 7, 12 and 16: *g* 7 = general cognitive ability age 7; *g* 12 = general cognitive ability age 12; *g* 16 = general cognitive ability age 16; F = female; M = male; *standardization was required to form composite.

Table S3. Study cohorts included in the *EduYears* GWAS summary statistics

Study	Full name	Sampling	Country	Sample size	Reference	Birth year (Mean/Range)	Fraction female
ACPRC	Manchester Studies of Cognitive Ageing	Population-based	England	1713	1	1923 (1903-1948)	0.71
AGES	Age, Gene/ Environment Susceptibility– Reykjavik Study	Population-based	Iceland	3212	2	1927 (1908-1936)	0.58
ALSPAC	Avon Longitudinal Study of Parents and Children	Population-based birth cohort	England	2877	3	1959 (1948-1963)	1.00
ASPS	Austrian Stroke Prevention Study	Population-based	Austria	777	4,5	1932 (1909-1949)	0.57
BASE-II	Berlin Aging Study II	Population-based	Germany	1619	6	1948 (1925-1983)	0.52
CoLaus	Cohorte Lausannoise	Population-based	Switzerland	3269	7	1950 (1928-1970)	0.53
COPSAC2000	Copenhagen Studies on Asthma in Childhood 2000	Case-control birth cohort	Germany	318	8	1966 (1964-1969)	0.47
CROATIA-Korčula	Croatia Korčula	Population-based (Isolate)	Croatia	842	9	1950 (1909-1977)	0.64
deCODE	deCODE genetics	Population-based	Iceland	46758	10	1945 (1894-1983)	0.57
DHS	Dortmund Health Study	Population-based	Germany	953	11	1949 (1929-1974)	0.53
DIL	Wellcome Trust Diabetes and Inflammation Laboratory	Population-based	England	2578	12	1958 (1958-1958)	0.52
EGCUT1	Estonian Genome Center, University of Tartu	Population-based	Estonia	5597	13	1950 (1905-1980)	0.55
EGCUT2	Same as above	Population-based	Estonia	1328	13	1957 (1911-1979)	0.53
EGCUT3	Same as above	Population-based	Estonia	2047	13	1966 (1930-1982)	0.73
ERF	Erasmus Rucphen Family Study	Family-based	Netherlands	2433	14,15	1952 (1914-1974)	0.55
FamHS	Family Heart Study	Family-based	USA	3483	16,17	1941 (1900-1965)	0.53
FINRISK	The National FINRISK Study	Case-control (Cardiovascular health)	Finland	1685	18	1946 (1923-1977)	0.46
FTC	Finnish Twin Cohort	Family-based	Finland	2418	19	1945 (1910-1972)	0.56
GOYA	Genetics of Overweight Young Adults	Case-control (Obesity)	Denmark	1459	20	1947 (1944-1954)	0.00
GRAPHIC	Genetic Regulation of Arterial Pressure in Humans	Population-based	England	727	21	1951 (1942-1965)	0.53
GS	Generation Scotland	Population-based	Scotland	8776	22	1955 (1909-1981)	0.59
H2000 Cases	Health 2000	Case-control (Metabolic syndrome)	Finland	797	23	1949 (1924-1970)	0.50
H2000 Controls	Same as above	Case-control (Metabolic syndrome)	Finland	819	23	1949 (1924-1969)	0.52
HBCS	Helsinki Birth Cohort Study	Population-based birth cohort	Finland	1617	24	1941 (1934-1944)	0.57
HCS	Hunter Community Study	Population-based	Australia	1946	25	1940 (1920-1951)	0.49
HNRS (CorexB)	Heinz Nixdorf Recall Study	Population-based	Germany	1401	25	1942 (1926-1955)	0.50

HNRS (Oexpr)	Same as above	Same as above	Germany	1347	25	1942 (1926-1955)	0.50
HNRS (Omni1)	Same as above	Same as above	Germany	778	26	1942 (1927-1955)	0.52
HRS	Health and Retirement Study	Population-based	USA	9963	27	1940 (1900-1979)	0.42
Hypergenes	Hypergenes	Case-control	Italy/ UK/ Belgium	815	28	1945 (1914-1971)	0.46
INGI-CARL	Italian Network of Genetic Isolates - Carlantino	Population-based (Isolate)	Italy	947	28	1946 (1910-1975)	0.58
INGI-FVG	Italian Network of Genetic Isolates - Friuli Venezia Giulia	Population-based (Isolate)	Italy	943	29	1951 (1917-1978)	0.60
KORA S3	Kooperative Gesundheitsforschung in der Region Augsburg	Population-based	Germany	2655	29	1945 (1920-1964)	0.51
KORA S4	Same as above	Population-based	Germany	2721	30	1949 (1926-1970)	0.51
LBC1921	Lothian Birth Cohort 1921	Population-based birth cohort	Scotland	515	31	1921 (1921-1921)	0.58
LBC1936	Lothian Birth Cohort 1936	Population-based birth cohort	Scotland	1003	32	1936 (1936-1936)	0.49
LifeLines	The LifeLines Cohort Study	Population-based	Netherlands	12539	33	1960 (1921-1980)	0.58
MCTFR	Minnesota Center for Twin and Family Research	Family-based, but only founders used.	USA	3819	34	1953 (1926-1974)	0.54
MGS	Molecular Genetics of Schizophrenia	Population-based	USA	2313	35,36	1951 (1914-1976)	0.50
MoBa	Mother and Child Cohort of NIPH	Population-based (Nested case-control)	Norway	622	37,38	1971 (1966-1976)	1.00
NBS	Nijmegen Biomedical Study	Population-based	Netherlands	1808	39	1941 (1923-1972)	0.50
NESDA	Netherlands Study of Depression and Anxiety	Case-control (Mental health)	Netherlands	1820	40	1958 (1939-1977)	0.64
NFBC66	Northern Finland Birth Cohort 1966	Population-based	Finland	5297	41,42	1966 (1966-1966)	0.52
NTR	Netherlands Twin Register	Family-based	Netherlands	5246	43	1958 (1917-1989)	0.64
OGP	Ogliastra Genetic Park	Population-based	Italy	370	44	1950 (1916-1976)	0.00
OGP-Talana	Ogliastra Genetic Park-Talana	Population-based (Isolate)	Italy	544	44	1949 (1910-1977)	0.59
ORCADES	Orkney Complex Disease Study	Population-based (Isolate)	Scotland	1828	45	1952 (1914-1979)	0.60
PREVEND	Prevention of Renal and Vascular End-stage Disease	Population-based	Netherlands	3578	46	1948 (1923-1968)	0.48
QIMR	Queensland Institute of Medical Research	Family-based	Australia	8006	47	1956 (1900-1984)	0.59
RS-I	Rotterdam Study Baseline	Population-based	Netherlands	6108	48,49	1922 (1893-1938)	0.60
RS-II	Rotterdam Study Extension of Baseline	Same as above	Netherlands	1667	48,49	1935 (1906-1944)	0.52
RS-III	Rotterdam Study Young	Same as above	Netherlands	3040	48,49	1950 (1910-1960)	0.56
Rush-MAP	Rush University Medical Center - Memory and Aging Project	Community-based	USA	887	50	1921 (1901-1948)	0.72
Rush-ROS	Rush University Medical Center - Religious Orders Study	Community-based	USA	808	51	1921 (1896-1946)	0.66
SardiNIA	SardiNIA Study of Aging	Family-based	Italy	5616	52	1955 (1901-1983)	0.58

SHIP	Study of Health in Pomerania	Population-based	Germany	3556	52	1945 (1918-1971)	0.50
SHIP-TREND	Study of Health in Pomerania	Population-based	Germany	901	53	1956 (1928-1980)	0.57
STR – Salty	Swedish Twin Registry	Family-based	Sweden	4832	52	1951 (1943-1958)	0.52
STR – Twingene	Swedish Twin Registry	Family-based	Sweden	9553	54	1941 (1916-1958)	0.53
THISEAS	The Hellenic Study of Interactions between SNPs & Eating in Atherosclerosis Susceptibility	Case-control	Greece	829	55	1950 (1909-1979)	0.33
TwinsUK	St Thomas' UK Adult Twin Registry	Population-based	England	4012	56	1949 (1919-1978)	1.00
WTCCC58C	1958 British Birth Cohort	Population-based	England	2804	57	1958 (1958-1958)	0.48
YFS	The Cardiovascular Risk in Young Finns Study	Population-based	Finland	2029	58	1969 (1962-1977)	0.55
UKB	UK Biobank	Population-based	UK	111349	59	1951 (1934-1970)	0.53

Note. Adapted from Okbay et al., *Nature*, in press[60]

Table S4. Numbers of SNPs per threshold used in creating *EduYears* GPS

pT	Number of SNPs
0.001	2,162
0.01	7,303
0.05	19,415
0.055[1]	20,582
0.063[2,3]	22,384
0.067[4,5]	23,316
0.072[6]	24,377
0.1	30,086
0.2	46,636
0.3	60,012
0.362[7]	67,265
0.364[8]	67,470
0.4	71,382
0.5	81,149

Note. pT = *P*-value threshold; [1]'Best-fit' GPS for *g* 16; [2]'Best-fit' GPS for *g* 7; [3]'Best-fit' GPS for *g* composite; [4]'Best-fit' GPS for *g* 12; [5]'Best-fit' GPS for EA 16; [6]'Best-fit' GPS for family SES; [7]'Best-fit' GPS for EA 12; [8]'Best-fit' GPS for EA 7.

Table S5. Testing significance of differences of the correlations between *EduYears* GPS and educational achievement at ages 7, 12 and 16.

	z	*P*-value
*r*GPS-EA7 - *r*GPS-EA12	1.65	0.10
*r*GPS-EA12 - *r*GPS-EA16	3.88	< 0.001
*r*GPS-EA7 - *r*GPS-EA16	6.29	< 0.001

Note. GPS = *EduYears* GPS; EA 7 = educational achievement age 7; EA 12 = educational achievement age 12; EA 16 = educational achievement age 16; *r* = Pearson's correlation.

Table S6. Correlations among all variables

	EA 7	EA 12	EA 16	*g* comp	*g* 7	*g* 12	*g* 16	SES	GPS
EA7	1								
EA12	0.60***	1							
EA16	0.59***	0.66***	1						
g	0.50***	0.53***	0.61***	1					
g 7	0.45***	0.41***	0.46***	0.76***	1				
g 12	0.44***	0.48***	0.53***	0.83***	0.45***	1			
g 16	0.41***	0.48***	0.58***	0.83***	0.42***	0.59***	1		
SES	0.33***	0.35***	0.46***	0.37***	0.33***	0.34***	0.33***	1	
GPS	0.17***	0.21***	0.30***	0.19***	0.15***	0.19***	0.20***	0.27***	1

Note. EA 7 = educational achievement age 7; EA 12 = educational achievement age 12; EA 16 = educational achievement age 16; *g* comp = general cognitive ability composite; *g* 7 = general cognitive ability age 7; *g* 12 = general cognitive ability age 12; *g* 16 = general cognitive ability age 16; SES = family socioeconomic status; GPS = genome-wide polygenic score; ***P < 0.001; *unique 'best-fit' GPS was used for each respective trait, see methods section for details.

Table S7. Predicting educational achievement at age 7, 12 and 16 from *g, SES* and *EduYears* GPS: Multiple Regression Analysis

	Dependent variable:		
	EA 7	EA 12	EA 16
g	**0.44***** (0.40 - 0.48)	**0.46***** (0.38 - 0.48)	**0.50***** (0.45 - 0.52)
SES	**0.14***** (0.09 - 0.19)	**0.17**** (0.11 - 0.22)	**0.25**** (0.22 - 0.30)
GPS	0.03 (-0.01 - 0.07)	0.07* (-0.01 - 0.11)	**0.11***** (0.07 - 0.14)
N	1,738	1,035	1,763
R^2	0.27	0.32	0.45
Adjusted R^2	0.27	0.31	0.45
Residual Std. Error	0.85 (df = 1,734)	0.80 (df = 1,031)	0.72 (df = 1,759)
F Statistic	212.50*** (df = 3; 1,734)	158.81*** (df = 3; 1,031)	474.64*** (df = 3; 1,759)

Note. Standardized coefficients are presented; 95% Confidence Intervals in brackets; *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; unique 'best-fit' GPS was used for each respective trait.

Table S8. Testing septile extreme group differences at age 7, 12 and 16 for *EduYears* GPS: Analysis of Variance

	N	*lower extreme Mean (SE)*	*upper extreme Mean (SE)*	*F*	*d*
EA 7	1,139	-0.26(0.04)	0.25(0.04)	**72.84*****	0.51
EA 12	754	-0.32(0.05)	0.31(0.05)	**74.46*****	0.63
EA 16	1,127	-0.47(0.04)	0.43(0.04)	**284.88*****	0.90

Note. Lower and upper extreme represent lowest and highest GPS septiles; *d* = Cohen's *d*; ****P* < 0.001; unique 'best-fit' GPS was used for each respective trait.

Table S9. Testing quintile extreme group differences at age 7, 12 and 16 for *EduYears* GPS: Analysis of Variance

	N	*lower extreme Mean (SE)*	*upper extreme Mean (SE)*	*F*	*d*
EA 7	1,585	-0.24(0.03)	0.21(0.03)	**82.01*****	0.45
EA 12	1,054	-0.28(0.04)	0.26(0.04)	**80.31*****	0.54
EA 16	1,703	-0.41(0.03)	0.40(0.03)	**319.26*****	0.81

Note. Lower and upper extreme represent lowest and highest GPS quintiles; *d* = Cohen's *d*; ****P* < 0.001; unique 'best-fit' GPS was used for each respective trait.

Table S10. Testing for GPSxSES interaction at age 7, 12 and 16: Multiple Regression Analysis

	Dependent variable:		
	EA 7	EA 12	EA 16
GPS	**0.09***** (0.05 - 0.12)	**0.13***** (0.09 - 0.17)	**0.19***** (0.16 - 0.22)
SES	**0.31***** (0.28 - 0.35)	**0.31***** (0.27 - 0.35)	**0.41***** (0.40 - 0.46)
GPS*SES	-0.01 (-0.04 - 0.02)	0.02 (-0.01 - 0.06)	-0.02 (-0.05 - 0.01)
N	3,848	2,341	3,804
R^2	0.12	0.14	0.25
Adjusted R^2	0.12	0.14	0.25
Residual Std. Error	0.94 (df = 3,844)	0.92 (df = 2,337)	0.86 (df = 3,800)
F Statistic	168.85*** (df = 3; 3,844)	124.23*** (df = 3; 2,337)	415.80*** (df = 3; 3,800)

Note. Standardized coefficients are presented; 95% Confidence Intervals in brackets; *P < 0.05; **P < 0.01; ***P < 0.001; unique 'best-fit' GPS was used for each respective trait.

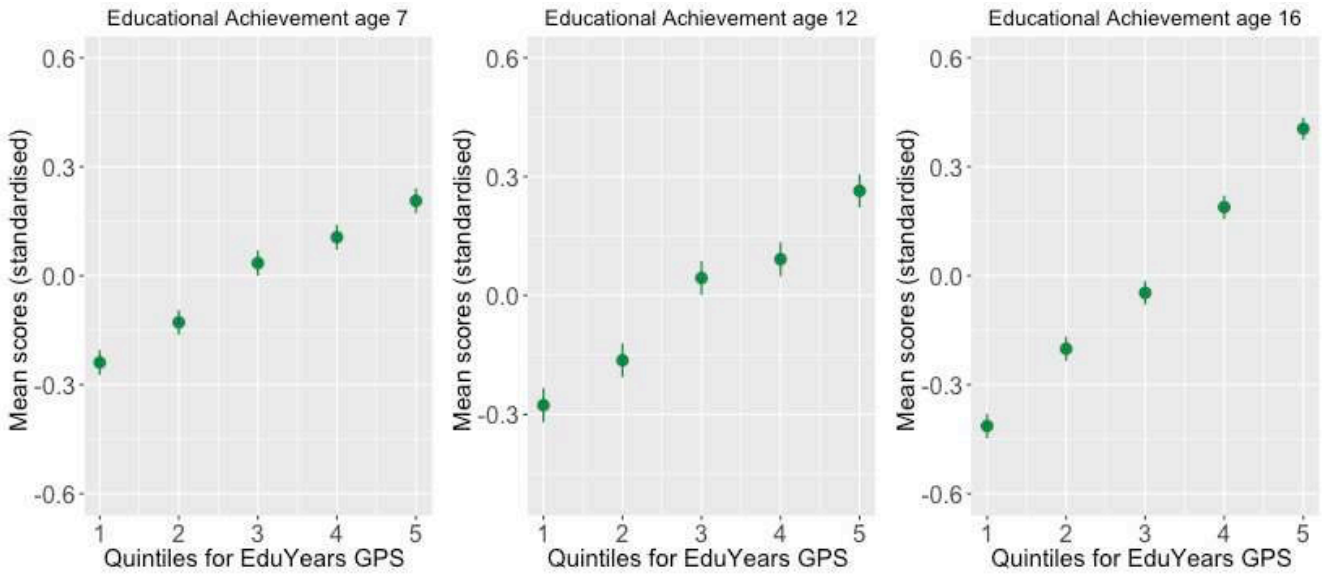
Table S11. Testing GPSxSES interaction for general cognitive ability: Multiple Regression Analysis

	Dependent variable:
	g
GPS	**0.11***** (0.07 - 0.16)
SES	**0.37***** (0.32 - 0.41)
GPS*SES	-0.03 (-0.07 - 0.01)
N	1,956
R^2	0.15
Adjusted R^2	0.15
Residual Std. Error	0.92 (df = 1,952)
F Statistic	118.32*** (df = 3; 1,952)

Note. Standardized coefficients are presented; 95% Confidence Intervals in brackets; *P < 0.05; **P < 0.01; ***P < 0.001; unique 'best-fit' GPS for g was used.

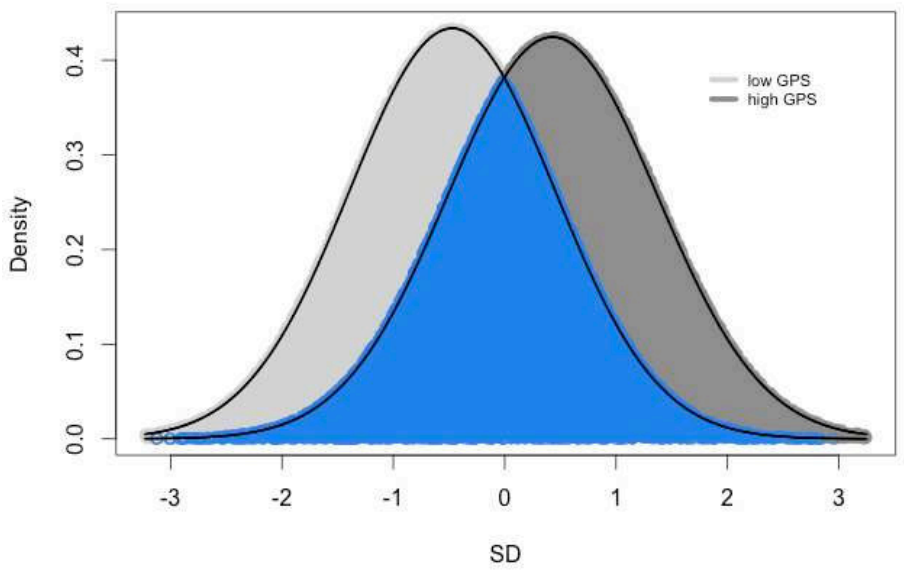
Supplementary Figures

Figure S1. Standardized means and standard errors for educational achievement at age 7, 12 and 16 by genome-wide polygenic score (GPS) quintile.



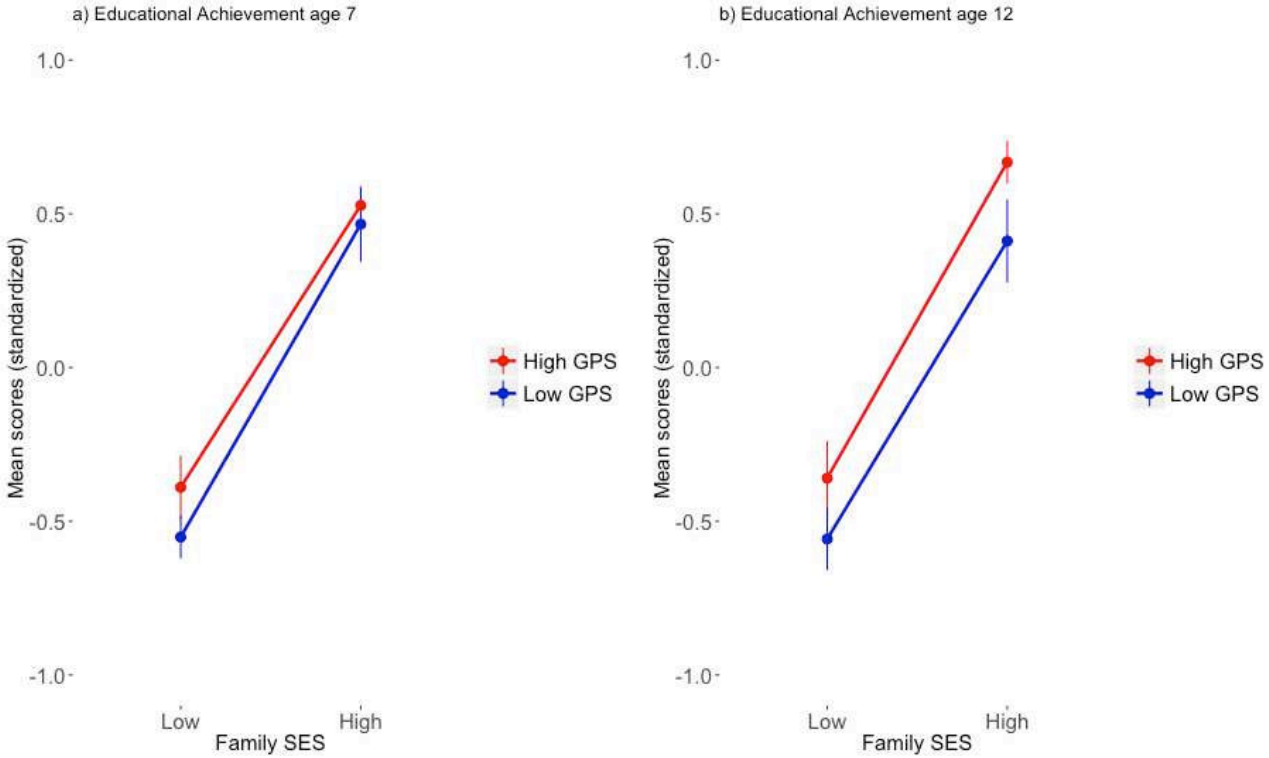
Note. *EduYears* GPS were rescored as quintiles (1=lowest, 5=highest); unique 'best-fit' GPS was used for each respective trait.

Figure S2. Probability density functions of standardized educational achievement scores age 16 for the lowest *EduYears* GPS septile (low GPS) and highest *EduYears* GPS septile (high GPS).



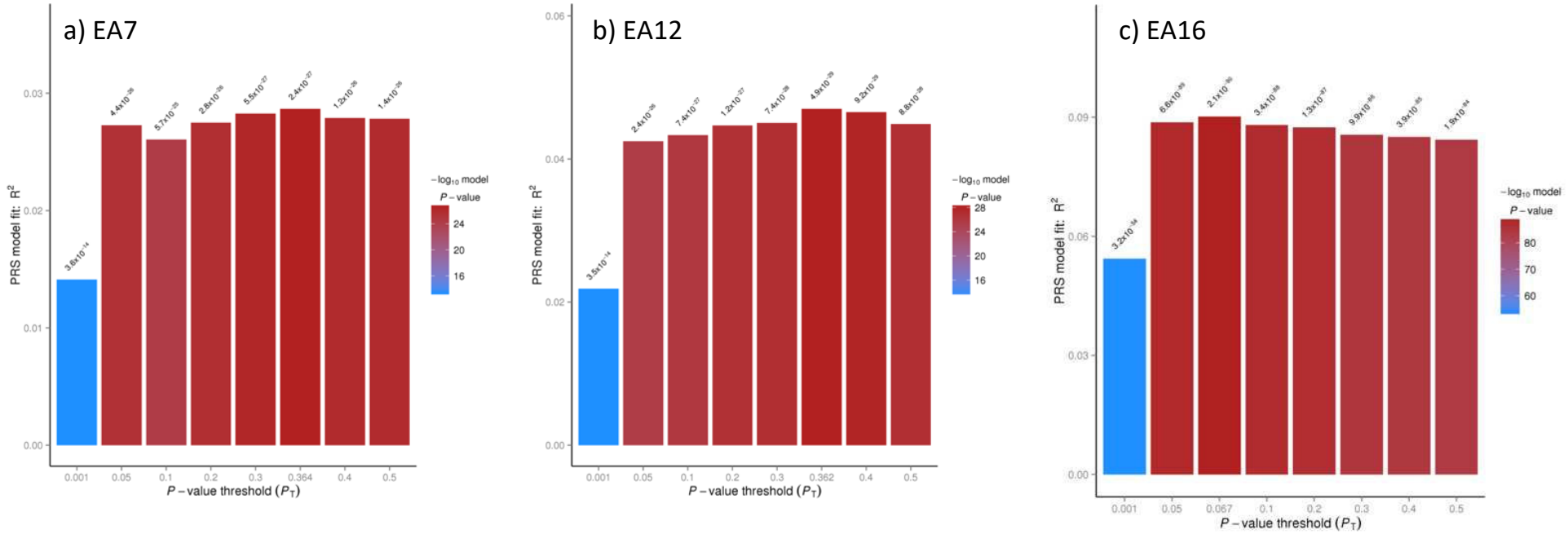
Note. SD = standard deviation. The area in blue indicates the overlap between the distributions.

Figure S3. Testing for GxE interaction



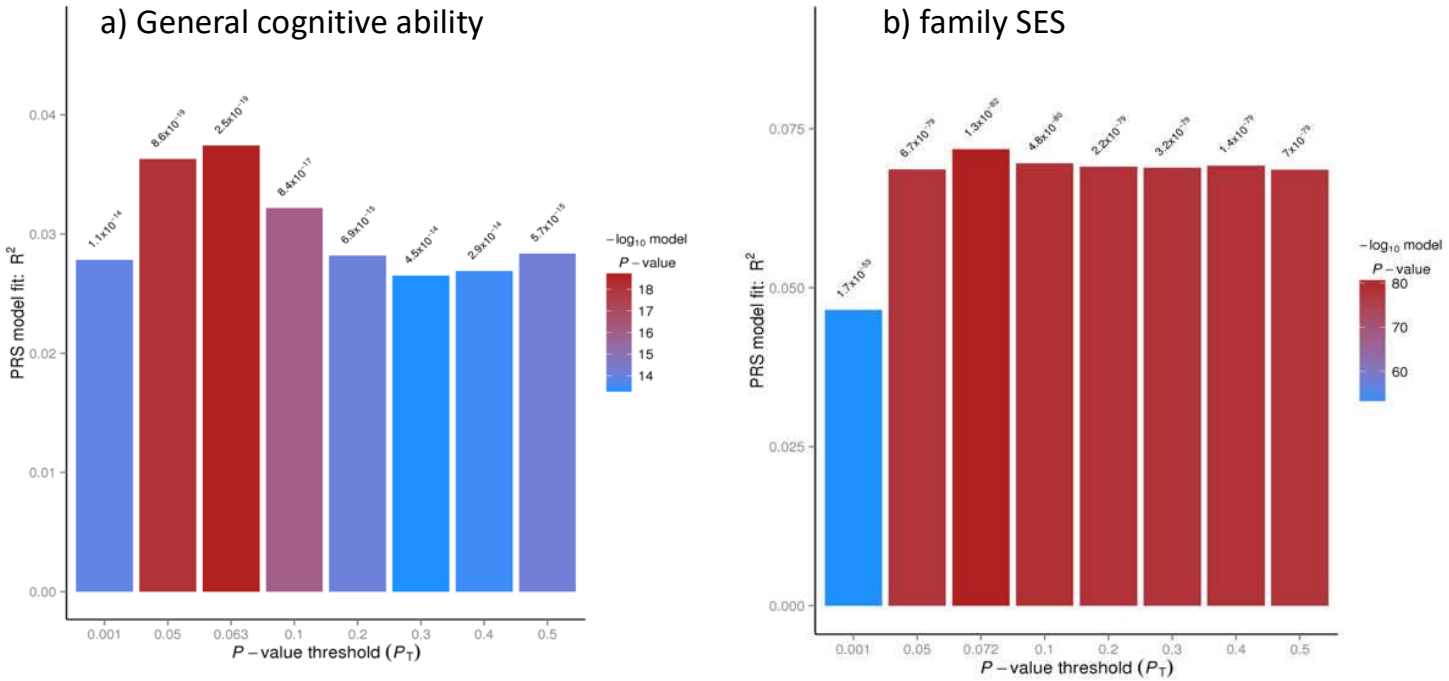
Note. Standardized educational achievement mean scores and error bars at a) age 7 and b) age 12 by *EduYears* GPS and family SES for individuals scoring in the highest and lowest quintiles of the distribution for both variables. No interaction effect was found at either age 7 ($F(1,617) = 0.29$, $P = 0.59$) or age 12 ($F(1,362) = 0.06$, $P = 0.80$); unique 'best-fit' GPS was used for both traits.

Figure S4. *EduYears* Genome-wide polygenic scores (GPS) explaining variance (R^2) in educational achievement



Note. a) age 7, b) age 12 and c) age 16. Different significance thresholds were used to include SNPs related to years of education ranging from 0.001 to 0.50. Using high-resolution scoring implemented in the PRSice software[61], a series of regression analyses determined the most predictive threshold to compute a GPS for educational achievement at all ages. The uncorrected P-values adjacent to each bar represent the statistical significance of the association between *EduYear*s GPS and the respective trait; see Supplementary Table S10 for the 'best-fit' GPS *P*-value threshold for educational achievement at age 7, 12 and 16.

Figure S5. *EduYears* Genome-wide polygenic scores (GPS) explaining variance (R^2)



Note. a) general cognitive ability and b) family SES. Different significance thresholds were used to include SNPs related to years of education ranging from 0.001 to 0.50. Using high-resolution scoring implemented in the PRSice software[61], a series of regression analyses determined the most predictive threshold to compute a GPS for educational achievement at all ages. The uncorrected P-values adjacent to each bar represent the statistical significance of the association between *EduYears* GPS and the respective trait; see Supplementary Table S10 for the 'best-fit' GPS *P*-value threshold for general cognitive ability and family SES.

Supplementary Methods

Methods S1. According to the AVENGEME software[62], a GPS constructed on the basis of a GWA discovery sample size of 328,918 in our target sample including 4,301 individuals (based on the sample size of the educational achievement measure at age 16) has more than 80% power to explain 0.2% of the phenotypic variance under the following circumstances; number of independent SNPs in the GPS = 50,000; proportion of total variance explained by genetic effects in discovery sample = 4%; covariance between genetic effect sizes in the discovery and target sample = 2%; proportion of SNPs with no effects on discovery trait = 99%; Range of P-values from GWA study summary statistics = 0 – 1.

References

1. Rabbitt, P. M. A. *et al.* The University of Manchester Longitudinal Study of Cognition in Normal Healthy Old Age, 1983 through 2003. *Aging Neuropsychol. Cogn.* **11,** 245–279 (2004).

2. Harris, T. B. *et al.* Age, Gene/Environment Susceptibility-Reykjavik Study: multidisciplinary applied phenomics. *Am. J. Epidemiol.* **165,** 1076–1087 (2007).

3. Fraser, A. *et al.* Cohort Profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. *Int. J. Epidemiol.* **42,** 97–110 (2013).

4. Schmidt, R. *et al.* Assessment of cerebrovascular risk profiles in healthy persons: definition of research goals and the Austrian Stroke Prevention Study (ASPS). *Neuroepidemiology* **13,** 308–313 (1994).

5. Schmidt, R., Fazekas, F., Kapeller, P., Schmidt, H. & Hartung, H. P. MRI white matter hyperintensities: three-year follow-up of the Austrian Stroke Prevention Study. *Neurology* **53,** 132–139 (1999).

6. Bertram, L. *et al.* Cohort profile: The Berlin Aging Study II (BASE-II). *Int. J. Epidemiol.* **43,** 703–712 (2014).

7. Firmann, M. *et al.* The CoLaus study: a population-based study to investigate the epidemiology and genetic determinants of cardiovascular risk factors and metabolic syndrome. *BMC Cardiovasc. Disord.* **8,** 1 (2008).

8. Bisgaard, H. The Copenhagen Prospective Study on Asthma in Childhood (COPSAC): design, rationale, and baseline data from a longitudinal birth cohort study. *Ann. Allergy Asthma Immunol. Off. Publ. Am. Coll. Allergy Asthma Immunol.* **93,** 381–389 (2004).

9. Rudan, I. *et al.* '10 001 Dalmatians:' Croatia Launches Its National Biobank. *Croat. Med. J.* **50,** 4–6 (2009).

10. Styrkarsdottir, U. *et al.* Nonsense mutation in the LGR4 gene is associated with several human diseases and other traits. *Nature* **497,** 517–520 (2013).

11. Pfaffenrath, V. *et al.* Regional variations in the prevalence of migraine and tension-type headache applying the new IHS criteria: the German DMKG Headache Study. *Cephalalgia Int. J. Headache* **29,** 48–57 (2009).

12. Strachan, D. P. *et al.* Lifecourse influences on health among British adults: effects of region of residence in childhood and adulthood. *Int. J. Epidemiol.* **36,** 522–531 (2007).

13. Nelis, M. *et al.* Genetic Structure of Europeans: A View from the North–East. *PLOS ONE* **4,** e5472 (2009).

14. Sayed-Tabatabaei, F. A. *et al.* Heritability of the function and structure of the arterial wall: findings of the Erasmus Rucphen Family (ERF) study. *Stroke J. Cereb. Circ.* **36,** 2351–2356 (2005).

15. Sleegers, K. *et al.* Cerebrovascular risk factors do not contribute to genetic variance of cognitive function: the ERF study. *Neurobiol. Aging* **28,** 735–741 (2007).

16. Higgins, M. *et al.* NHLBI Family Heart Study: objectives and design. *Am. J. Epidemiol.* **143,** 1219–1228 (1996).

17. O'Donnell, C. J. *et al.* Genome-wide association study for coronary artery calcification with follow-up in myocardial infarction. *Circulation* **124,** 2855–2864 (2011).

18. Vartiainen, E. *et al.* Thirty-five-year trends in cardiovascular risk factors in Finland. *Int. J. Epidemiol.* **39,** 504–518 (2010).

19. Kaprio, J. The Finnish Twin Cohort Study: An Update. *Twin Res. Hum. Genet.* **16,** 157–162 (2013).

20. Paternoster, L. *et al.* Genome-Wide Population-Based Association Study of Extremely Overweight Young Adults – The GOYA Study. *PLOS ONE* **6,** e24303 (2011).

21. Tobin, M. D. *et al.* Common variants in genes underlying monogenic hypertension and hypotension and blood pressure in the general population. *Hypertension* **51,** 1658–1664 (2008).

22. Smith, B. H. *et al.* Cohort Profile: Generation Scotland: Scottish Family Health Study (GS:SFHS). The study, its participants and their potential for genetic research on health and illness. *Int. J. Epidemiol.* **42,** 689–700 (2013).

23. Aromaa, A. *Health and functional capacity in Finland : Baseline results of the Health 2000 health examination survey*. (National Public Health Institute = Kansanterveyslaitos, 2004).

24. Barker, D. J. P., Osmond, C., Forsén, T. J., Kajantie, E. & Eriksson, J. G. Trajectories of growth among children who have coronary events as adults. *N. Engl. J. Med.* **353,** 1802–1809 (2005).

25. McEvoy, M. *et al.* Cohort profile: The Hunter Community Study. *Int. J. Epidemiol.* **39,** 1452–1463 (2010).

26. Mahabadi, A. A. *et al.* The Heinz Nixdorf Recall study and its potential impact on the adoption of atherosclerosis imaging in European primary prevention guidelines. *Curr. Atheroscler. Rep.* **13,** 367–372 (2011).

27. Sonnega, A. *et al.* Cohort Profile: the Health and Retirement Study (HRS). *Int. J. Epidemiol.* **43,** 576–585 (2014).

28. Salvi, E. *et al.* Genomewide association study using a high-density single nucleotide polymorphism array and case-control design identifies a novel essential hypertension

susceptibility locus in the promoter region of endothelial NO synthase. *Hypertension* **59,** 248–255 (2012).

29. Esko, T. *et al.* Genetic characterization of northeastern Italian population isolates in the context of broader European genetic diversity. *Eur. J. Hum. Genet. EJHG* **21,** 659–665 (2013).

30. Wichmann, H.-E., Gieger, C., Illig, T. & MONICA/KORA Study Group. KORA-gen--resource for population genetics, controls and a broad spectrum of disease phenotypes. *Gesundheitswesen Bundesverb. Ärzte Öffentl. Gesundheitsdienstes Ger.* **67 Suppl 1,** S26–30 (2005).

31. Deary, I. J., Whiteman, M. C., Starr, J. M., Whalley, L. J. & Fox, H. C. The impact of childhood intelligence on later life: following up the Scottish mental surveys of 1932 and 1947. *J. Pers. Soc. Psychol.* **86,** 130–147 (2004).

32. Deary, I. J. *et al.* The Lothian Birth Cohort 1936: a study to examine influences on cognitive ageing from age 11 to age 70 and beyond. *BMC Geriatr.* **7,** 28 (2007).

33. Stolk, R. P. *et al.* Universal risk factors for multifactorial diseases: LifeLines: a three-generation population-based study. *Eur. J. Epidemiol.* **23,** 67–74 (2008).

34. Miller, M. B. *et al.* The Minnesota Center for Twin and Family Research genome-wide association study. *Twin Res. Hum. Genet. Off. J. Int. Soc. Twin Stud.* **15,** 767–774 (2012).

35. Sanders, A. R. *et al.* No significant association of 14 candidate genes with schizophrenia in a large European ancestry sample: implications for psychiatric genetics. *Am. J. Psychiatry* **165,** 497–506 (2008).

36. Shi, J. *et al.* Common variants on chromosome 6p22.1 are associated with schizophrenia. *Nature* **460,** 753–757 (2009).

37. Magnus, P. *et al.* Cohort profile: the Norwegian Mother and Child Cohort Study (MoBa). *Int. J. Epidemiol.* **35,** 1146–1150 (2006).

38. Irgens, L. M. The Medical Birth Registry of Norway. Epidemiological research and surveillance throughout 30 years. *Acta Obstet. Gynecol. Scand.* **79,** 435–439 (2000).

39. Wetzels, J. F. M., Kiemeney, L. a. L. M., Swinkels, D. W., Willems, H. L. & den Heijer, M. Age- and gender-specific reference values of estimated GFR in Caucasians: the Nijmegen Biomedical Study. *Kidney Int.* **72,** 632–637 (2007).

40. Penninx, B. W. J. H. *et al.* The Netherlands Study of Depression and Anxiety (NESDA): rationale, objectives and methods. *Int. J. Methods Psychiatr. Res.* **17,** 121–140 (2008).

41. Sabatti, C. *et al.* Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. *Nat. Genet.* **41,** 35–46 (2009).

42. Rantakallio, P. Groups at risk in low birth weight infants and perinatal mortality. *Acta Paediatr. Scand.* **193,** Suppl 193:1+ (1969).

43. Boomsma, D. I. *et al.* Netherlands Twin Register: from twins to twin families. *Twin Res. Hum. Genet. Off. J. Int. Soc. Twin Stud.* **9,** 849–857 (2006).

44. Pistis, G. *et al.* High Differentiation among Eight Villages in a Secluded Area of Sardinia Revealed by Genome-Wide High Density SNPs Analysis. *PLOS ONE* **4,** e4654 (2009).

45. McQuillan, R. *et al.* Runs of homozygosity in European populations. *Am. J. Hum. Genet.* **83,** 359–372 (2008).

46. Hillege, H. L. *et al.* Urinary albumin excretion predicts cardiovascular and noncardiovascular mortality in general population. *Circulation* **106,** 1777–1782 (2002).

47. Martin, N. W. *et al.* Educational Attainment: A Genome Wide Association Study in 9538 Australians. *PLOS ONE* **6,** e20128 (2011).

48. Hofman, A. *et al.* The Rotterdam Study: 2012 objectives and design update. *Eur. J. Epidemiol.* **26,** 657–686 (2011).

49. Estrada, K. *et al.* GRIMP: a web- and grid-based tool for high-speed analysis of large-scale genome-wide association using imputed data. *Bioinforma. Oxf. Engl.* **25,** 2750–2752 (2009).

50. Bennett, D. A. *et al.* Overview and findings from the rush Memory and Aging Project. *Curr. Alzheimer Res.* **9,** 646–663 (2012).

51. Bennett, D. A., Schneider, J. A., Arvanitakis, Z. & Wilson, R. S. Overview and findings from the religious orders study. *Curr. Alzheimer Res.* **9,** 628–645 (2012).

52. Pilia, G. *et al.* Heritability of Cardiovascular and Personality Traits in 6,148 Sardinians. *PLOS Genet* **2,** e132 (2006).

53. Völzke, H. *et al.* Cohort profile: the study of health in Pomerania. *Int. J. Epidemiol.* **40,** 294–307 (2011).

54. Magnusson, P. K. E. *et al.* The Swedish Twin Registry: establishment of a biobank and other recent developments. *Twin Res. Hum. Genet. Off. J. Int. Soc. Twin Stud.* **16,** 317–329 (2013).

55. Theodoraki, E. V. *et al.* Fibrinogen beta variants confer protection against coronary artery disease in a Greek case-control study. *BMC Med. Genet.* **11,** 28 (2010).

56. Moayyeri, A., Hammond, C. J., Valdes, A. M. & Spector, T. D. Cohort Profile: TwinsUK and Healthy Ageing Twin Study. *Int. J. Epidemiol.* **42,** 76–85 (2013).

57. Power, C. & Elliott, J. Cohort profile: 1958 British birth cohort (National Child Development Study). *Int. J. Epidemiol.* **35,** 34–41 (2006).

58. Raitakari, O. T. *et al.* Cohort profile: the cardiovascular risk in Young Finns Study. *Int. J. Epidemiol.* **37,** 1220–1226 (2008).

59. Sudlow, C. *et al.* UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLOS Med* **12,** e1001779 (2015).

60. Okbay, A. *et al.* Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* (in press).
61. Euesden, J., Lewis, C. M. & O'Reilly, P. F. PRSice: Polygenic Risk Score software. *Bioinformatics* **31,** 1466–1468 (2015).
62. Palla, L. & Dudbridge, F. A Fast Method that Uses Polygenic Scores to Estimate the Variance Explained by Genome-wide Marker Panels and the Proportion of Variants Affecting a Trait. Am. J. Hum. Genet. 97, 250–259 (2015).

Appendix 2

Polygenic score for educational attainment captures DNA variants shared between personality traits and educational achievement

SUPPLEMENTARY METHODS

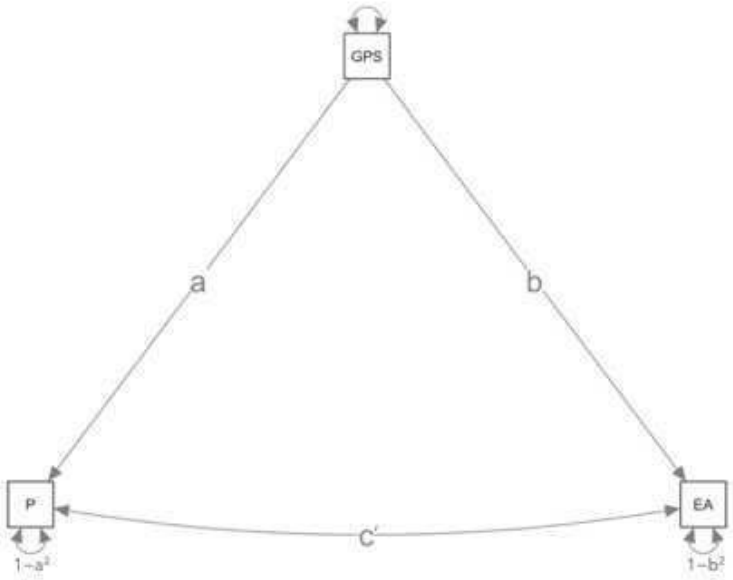
SUPPLEMENTARY TABLES

SUPPLEMENTARY FIGURES

Supplementary Methods

Methods S1 – Structural equation model of personality domains, GCSE results and polygenic scores

To test the extent to which the covariance between personality domains and GCSE results are explained by the polygenic scores, we used structural equation modelling. Because we assume causality from polygenic score (an aggregate score of DNA variants) to outcome variables, we applied the following model to our z-standardized variables:



Note. P = personality trait, EA = educational achievement defined as GCSE results, GPS = genome-wide polygenic score

Paths a, b and c are the beta effect size parameters. The GPS effect is described by the product of a and b, which is the pathway from the causal variable GPS to P, and GPS to EA. Path c' describes the residual relation between P and EA after accounting for the effects of the causal variable GPS in P and EA, respectively. The total effect can be derived by summing the effects of the residual and the indirect path, described as c' + ab. To calculate the proportion of the total effect that is explained by the causal variable GPS, the effect of the indirect path ab is divided by the total effect c' + ab

Supplementary Tables

Table S1 – Descriptive statistics of all variables for the full sample

		Mean (SD)			ANOVA of gender		ANOVA of age	
	N	Whole sample	Males	Females	*F*	R^2	*F*	R^2
Academic motivation Composite	**3079**	**0.00 (1.00)**	**-0.02 (0.97)**	**0.02 (1.02)**	-	-	-	-
PISA math self-efficacy	3078	17.53 (5.51)	18.94 (5.00)	16.55 (5.64)	99.66***	0.05	0.08	<0.01
PISA math interest	3079	2.53 (0.93)	2.65 (0.90)	2.45 (0.95)	22.9***	0.01	2.89	<0.01
PISA time spent on math	3050	4.48 (1.71)	4.48 (1.83)	4.48 (1.63)	0.00	<0.01	0.42	<0.01
Attitude towards key subjects	3078	2.54 (0.49)	2.52 (0.48)	2.56 (0.49)	2.5	<0.01	1.08	<0.01
Openness Composite	**2881**	**0.00 (1.00)**	**-0.02 (0.96)**	**0.01 (1.03)**	-	-	-	-
Academic self-concept	2837	3.55 (0.63)	3.65 (0.59)	3.48 (0.64)	35.53***	0.02	0.14	<0.01
Ambition	2837	3.89 (0.67)	3.9 (0.64)	3.88 (0.69)	0.55	<0.01	0.01	<0.01
Curiosity	3221	4.77 (0.9)	4.88 (0.88)	4.70 (0.91)	23.18***	0.01	0.06	<0.01
Hopefulness	3227	4.69 (0.72)	4.78 (0.69)	4.63 (0.73)	23.94***	0.01	1.96	<0.01
Openness	2803	3.58 (0.58)	3.53 (0.58)	3.62 (0.57)	9.11***	<0.01	1.78	<0.01
GRIT	2887	3.26 (0.59)	3.21 (0.57)	3.30 (0.61)	11.91***	0.01	0.91	<0.01
Optimism	2887	3.23 (0.71)	3.28 (0.69)	3.19 (0.72)	8.89***	<0.01	1.52	<0.01
Conscientiousness Composite	**2713**	**0.02 (0.99)**	**0.02 (0.98)**	**0.02 (0.99)**	-	-	-	-
SDQ Hyperactivity	6001	6.45 (2.30)	6.38 (2.30)	6.50 (2.30)	2.96	<0.01	2.30	<0.01
SWAN Hyperactivity	1313	4.78 (0.99)	4.8 (0.94)	4.77 (1.01)	0.19	<0.01	0.00	<0.01
SWAN Inattention	1313	4.64 (0.87)	4.62 (0.88)	4.66 (0.86)	0.31	<0.01	1.74	<0.01
Conscientiousness	2796	3.71 (0.60)	3.62 (0.600)	3.77 (0.59)	27.52***	0.01	0.00	<0.01

Agreeableness Composite	**6612**	**-0.01 (0.99)**	**-0.02 (1.01)**	**0.00 (0.98)**	-	-	-	-
Agreeableness	2798	3.66 (0.56)	3.53 (0.55)	3.75 (0.56)	74.16***	0.04	0.13	<0.01
SDQ Prosocial behavior	6001	7.14 (1.93)	6.52 (1.92)	7.63 (1.78)	357.53***	0.08	0.96	<0.01
School engagement	3068	3.00 (0.68)	2.98 (0.68)	3.02 (0.67)	2.16	<0.01	0.26	<0.01
SDQ Conduct scale	6000	8.39 (1.44)	8.29 (1.48)	8.46 (1.41)	13.52***	<0.01	3.09	<0.01
Gratitude	3229	5.8 (0.84)	5.65 (0.84)	5.91 (0.83)	51.36***	0.02	0.10	<0.01
Neuroticism Composite	**6008**	**-0.01 (1.01)**	**-0.02 (0.90)**	**0.00 (1.08)**	-	-	-	-
Cognitive Disorganization	5998	3.93 (2.86)	3.37 (2.73)	4.38 (2.88)	126.06***	0.03	1.26	<0.01
CASI anxiety	6004	8.07 (5.96)	6.20 (4.85)	9.53 (6.34)	332.69***	0.08	0.38	<0.01
MFQ	6003	3.59 (4.33)	2.56 (3.29)	4.39 (4.85)	183.36***	0.04	2.32	<0.01
Subjective happiness	5998	2.88 (0.94)	2.89 (0.91)	2.87 (0.97)	0.18	<0.01	3.80	<0.01
Life satisfaction	3224	2.40 (0.62)	2.38 (0.58)	2.41 (0.64)	1.76	<0.01	3.21	<0.01
Peer problems	6001	1.54 (1.51)	1.60 (1.53)	1.48 (1.49)	6.61*	<0.01	4.06*	<0.01
Neuroticism	2808	2.58 (0.68)	2.47 (0.64)	2.66 (0.69)	35.82***	0.02	6.87*	<0.01
Extraversion	**2807**	**0.00 (0.97)**	**-0.01 (0.97)**	**0.01 (0.97)**	**0.32**	**<0.01**	**0.69**	**<0.01**
GCSE	**8322**	**-0.01 (1.00)**	**-0.02 (1.02)**	**0.00 (0.99)**	**0.79**	**<0.01**	**1.71**	**<0.01**
general cognitive ability	**3939**	**0.01 (0.99)**	**0.00 (0.98)**	**0.01 (0.99)**	-	-	-	-
EduYears GPS	**10346**	**0.01 (1.00)**	**0.02 (1.01)**	**0.00 (0.99)**	**1.08**	**<0.01**	**0.16**	**<0.01**
Neuroticism GPS	**10346**	**0.01 (1.00)**	**0.01 (1.01)**	**0.01 (0.99)**	**0.02**	**<0.01**	**0.10**	**<0.01**
Wellbeing GPS	**10346**	**0.00 (1.00)**	**0.00 (0.99)**	**0.00 (1.01)**	**0.00**	**<0.01**	**0.01**	**<0.01**

Note: Means and standard deviations for individual measures are calculated based on raw data. Means and standard deviations for domains are calculated with z-standardised age and sex regressed data. Values of standard deviation are given in parentheses. $^{+}$ = standardization of the individual cognitive scales assessed at age 7, 12 and 16 was required to form this composite. N= sample size after exclusions. For DZ twin pairs, ANOVA performed on one randomly selected twin per pair to test the effect of sex and age. Results = F statistic; * = p<.05; ** = p<.01; R^2= proportion of variance explained.

Table S2 – Sensitivity analysis of missingness of personality/motivation composites on socio-economic status, general cognitive ability, and GCSE grades

GCSE					
	Present	Missing			
	M (SD)	M (SD)	*z*	*p*	R^2
Academic motivation	0.14 (0.97)	-0.09 (1.01)	8.01	6.63E-16	0.009
Openness Composite	0.15 (0.96)	-0.09 (1.01)	8.46	1.31E-17	0.010
Conscientiousness Composite	0.17 (0.98)	-0.09 (1.00)	8.69	1.73E-18	0.011
Agreeableness Composite	0.10 (0.98)	-0.30 (1.01)	12.73	5.77E-39	0.026
Neuroticism Composite	0.12 (0.98)	-0.26 (1.00)	13.07	6.06E-41	0.025
Extraversion	0.17 (0.95)	-0.09 (1.01)	9.23	1.00E-20	0.013

General cognitive ability					
	Present	Missing			
	M (SD)	M (SD)	*z*	*p*	R^2
Academic motivation	0.03 (0.98)	-0.05 (0.99)	1.96	0.05	0.001
Openness Composite	0.03 (0.99)	-0.05 (0.99)	2.09	0.04	0.001
Conscientiousness Composite	0.09 (0.97)	-0.11 (1.00)	5.16	1.98E-07	0.007
Agreeableness Composite	0.03 (0.98)	-0.16 (1.03)	3.22	1.24E-3	0.005
Neuroticism Composite	0.06 (0.98)	-0.17 (1.00)	5.26	1.18E-07	0.009
Extraversion	0.05 (0.97)	-0.08 (1.01)	3.41	6.16E-4	0.003

SES					
	Present	Missing			
	M (SD)	M (SD)	*z*	*p*	R^2
Academic motivation	0.2 (0.98)	0.04 (1.01)	5.91	3.09E-09	0.004
Openness Composite	0.21 (0.98)	0.04 (1.00)	6.23	3.96E-10	0.005
Conscientiousness Composite	0.22 (0.98)	0.04 (1.00)	6.26	3.20E-10	0.005
Agreeableness Composite	0.21 (0.98)	-0.14 (1.00)	13.47	7.88E-43	0.022
Neuroticism Composite	0.23 (0.98)	-0.12 (1.00)	13.85	3.02E-45	0.022
Extraversion	0.22 (0.98)	0.04 (1.01)	6.67	2.07E-11	0.006

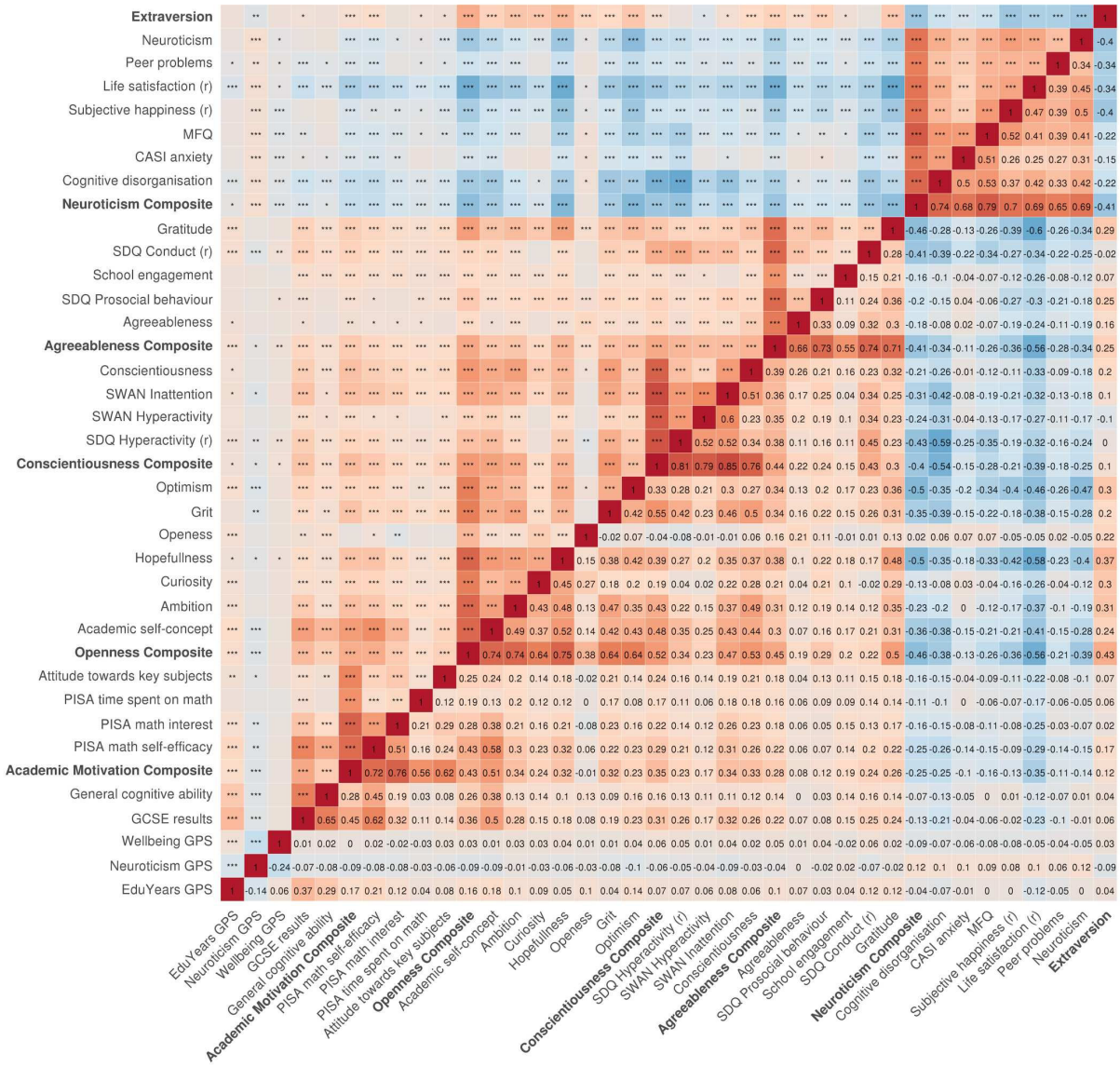
Note: SES = socio-economic status; GCSE = General Certificate for Secondary Education; present = data is non-missing for the respective composite; missing = data is missing for the respective composite. Analyses were performed on unrelated individuals only, where one twin per twin pair was randomly selected if data was available on dizygotic twin pairs. Degrees of freedom (df) for GCSE analyses = 5,600; df for general cognitive ability analyses = 2,676; df for SES analyses = 6,569.

Table S3 – Descriptive statistics of all variables for the sample of unrelated individuals

	N	*Mean*	*SD*	*Skew*	*Min*	*Max*
Academic motivation Composite	2084	0.00	1.00	-0.57	-4.4	2.49
Openness Composite	1958	0.00	1.00	-0.21	-4.48	3.12
Conscientiousness Composite	1838	0.02	0.99	-0.24	-3.52	2.6
Agreeableness Composite	4415	-0.01	0.99	-0.62	-4.56	2.62
Neuroticism Composite	4005	-0.01	1.01	0.89	-2.14	5.37
Extraversion	1909	0.00	0.97	-0.42	-3.86	2.26
GCSE	5602	-0.01	1.00	-0.43	-3.63	1.87
EduYears GPS	7026	0.01	1.00	0.08	-3.49	3.71
Neuroticism GPS	7026	0.01	1.00	0.04	-5.39	4.34
Wellbeing GPS	7026	0.00	1.00	0.01	-3.96	3.78

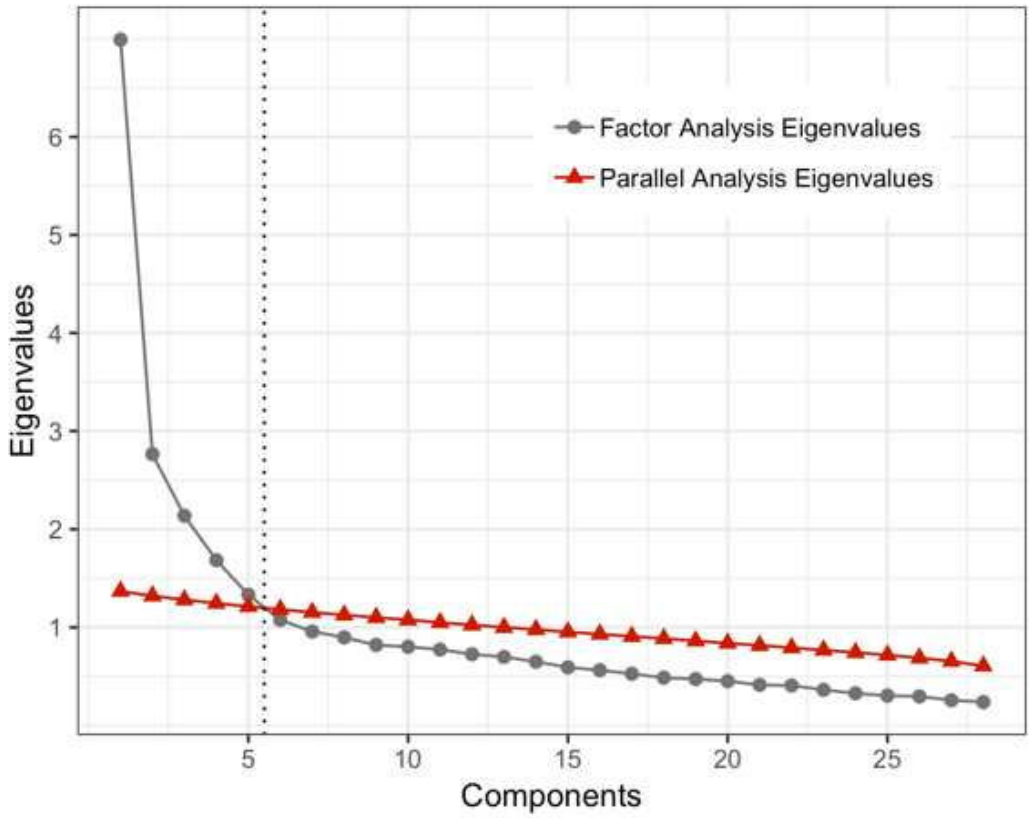
Supplementary Figures

Figure S1 – Correlations across all individual measures of personality and motivation, the personality/motivation composites and polygenic scores



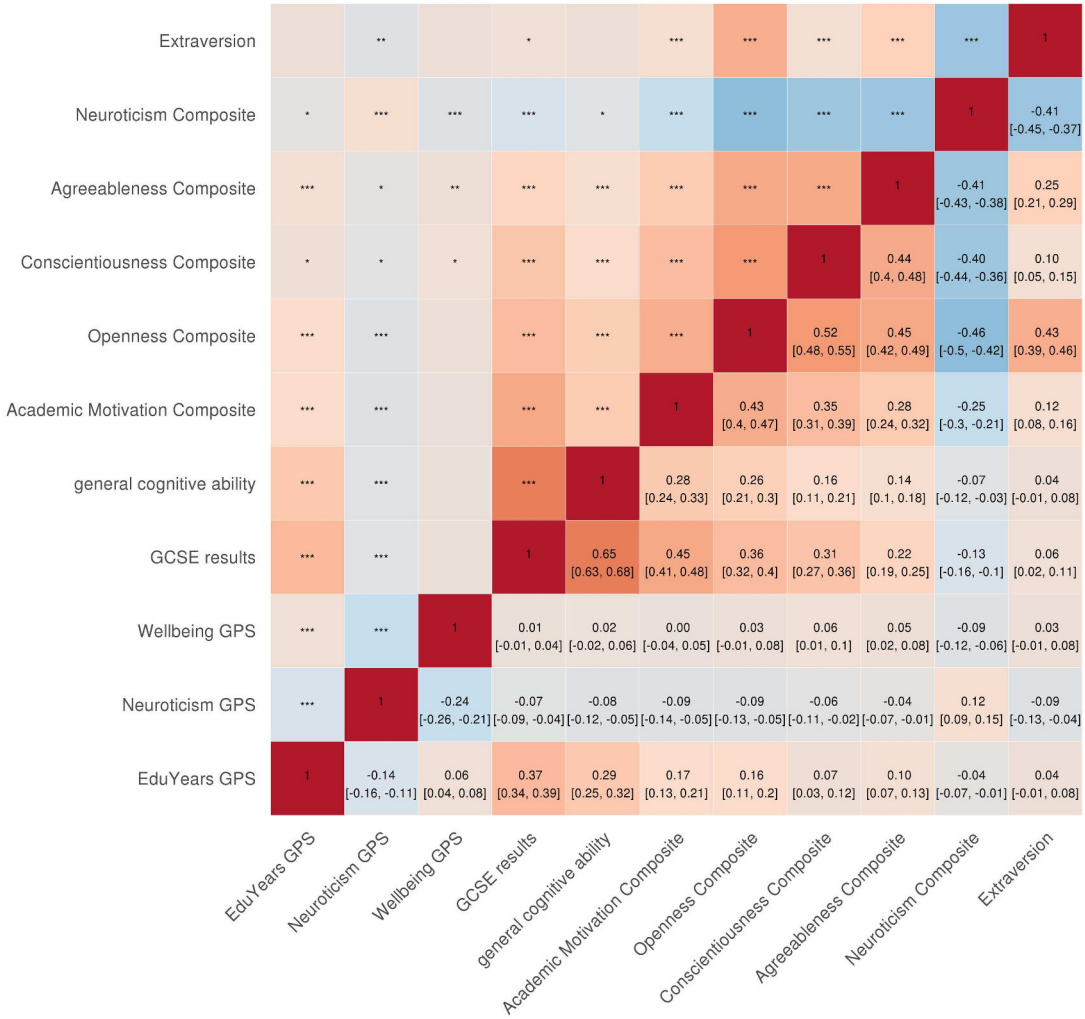
Note: (r.) = recoded so that higher scores were positive, i.e. less conduct problems. Variable labels in bold represent composites made up of the succeeding individual scales. * = p < 0.05; ** = p < 0.001; *** = p < 0.0001.

Figure S2 – Results from the parallel analysis.



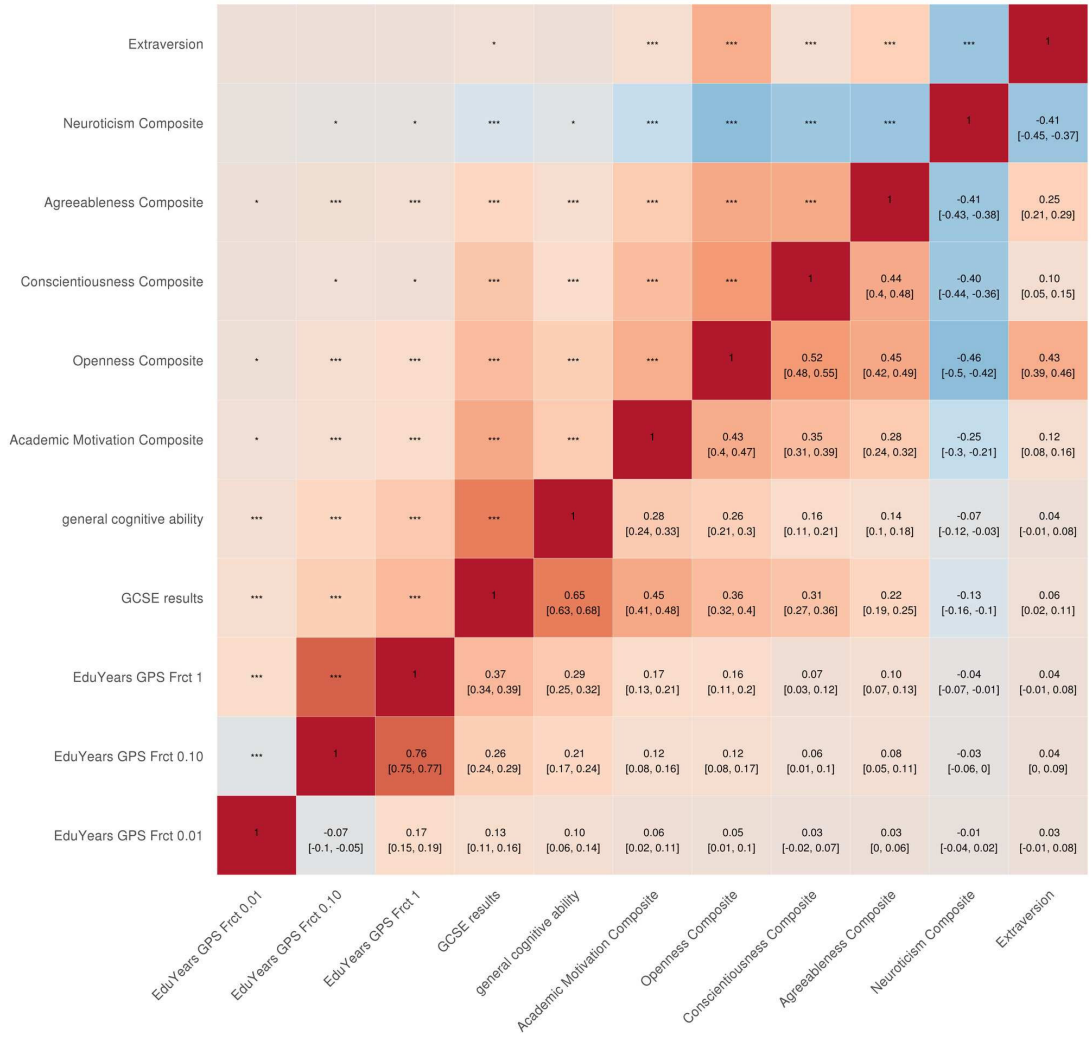
Note: The dotted line represents the point at which the factor eigenvalue in the study becomes smaller than the 95th percentile of the simulated eigenvalues from parallel analysis

Figure S3 – Correlations across all polygenic scores and personality/motivation domains



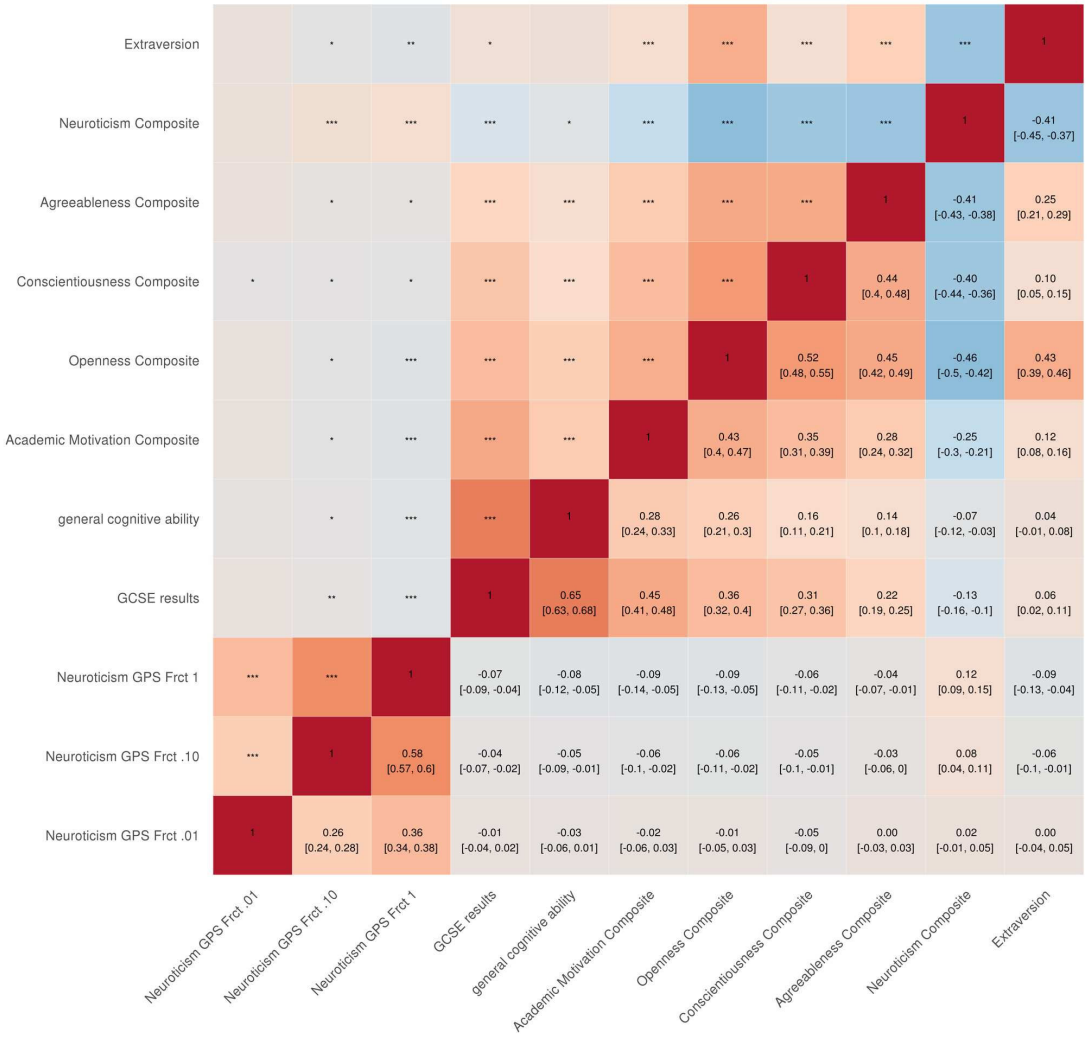
Note: 95% confidence intervals of the Pearson's correlation coefficients shown in square brackets. * = p < 0.05; ** = p < 0.001; *** = p < 0.0001.

Figure S4 – Correlations across all *EduYears* GPS thresholds and personality/motivation domains



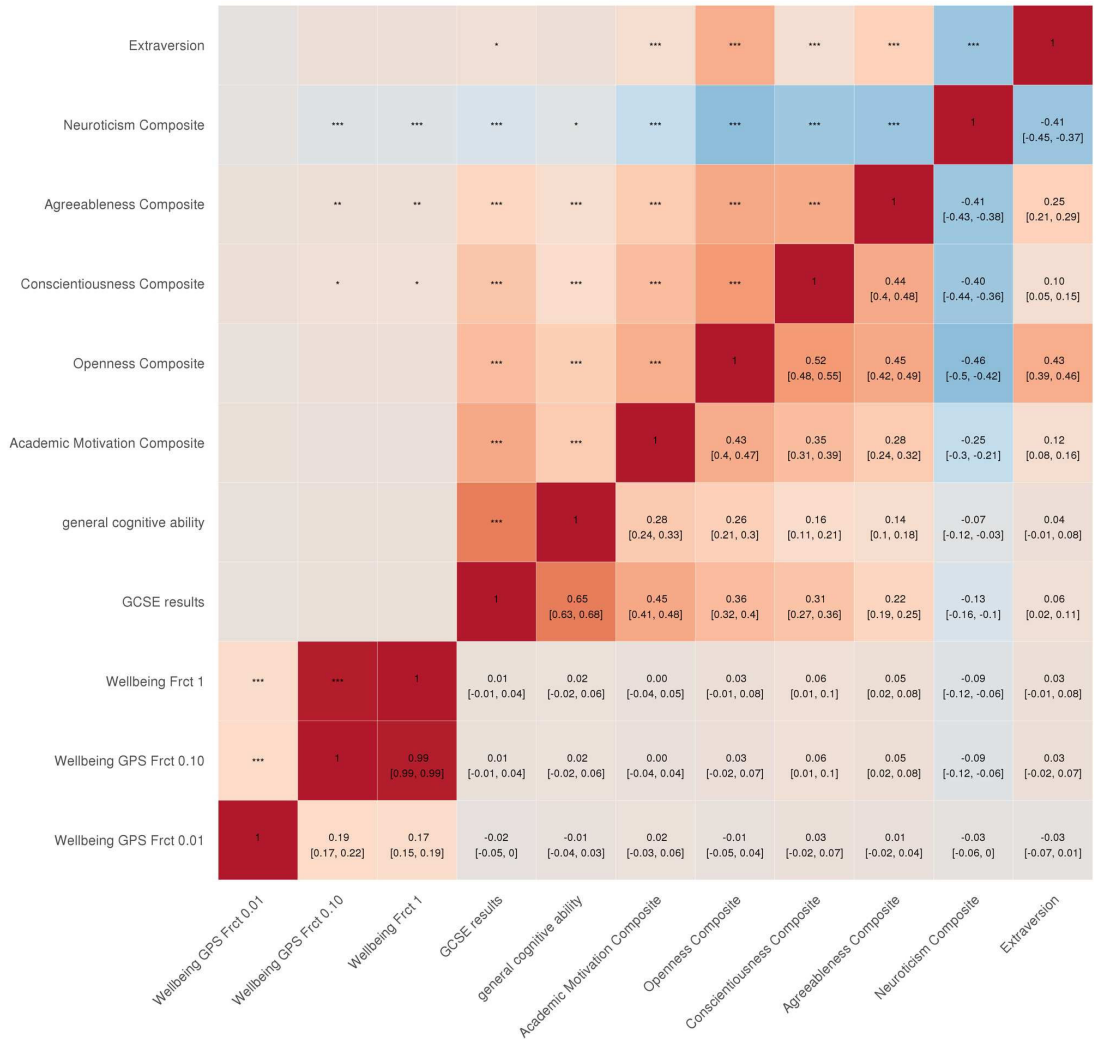
Note: 95% confidence intervals of the Pearson's correlation coefficients shown in square brackets. * = p < 0.05; ** = p < 0.001; *** = p < 0.0001.

Figure S5 – Correlations across all Neuroticism GPS thresholds and personality/motivation domains



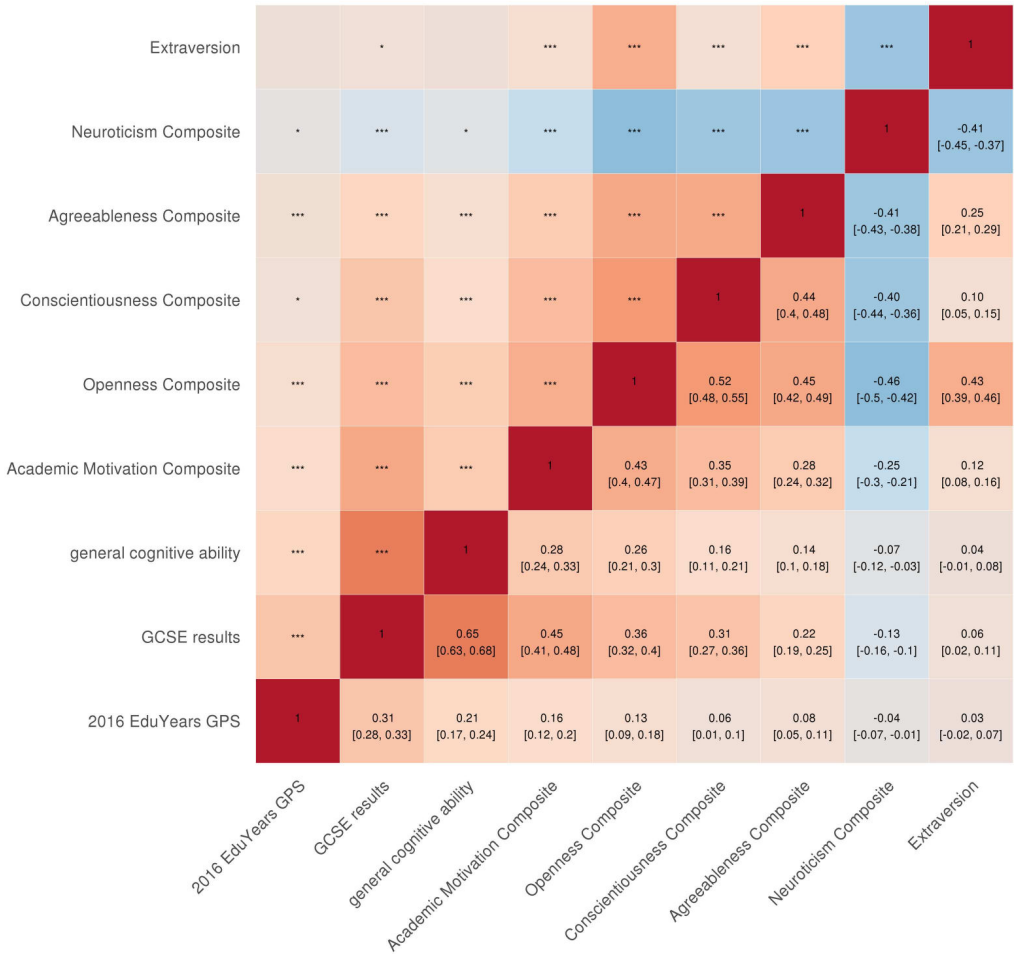
Note: 95% confidence intervals of the Pearson's correlation coefficients shown in square brackets. * = p < 0.05; ** = p < 0.001; *** = p < 0.0001.

Figure S6 – Correlations across all Wellbeing GPS thresholds and personality/motivation domains



Note: 95% confidence intervals of the Pearson's correlation coefficients shown in square brackets. * = p < 0.05; ** = p < 0.001; *** = p < 0.0001.

Figure S7 – Correlations between the 2016 *EduYears* GPS and outcome measures



Note: 95% confidence intervals of the Pearson's correlation coefficients shown in square brackets. * = p < 0.05; ** = p < 0.001; *** = p < 0.0001.

Appendix 3

A polygenic *p* factor for major psychiatric disorders

SUPPLEMENTARY TABLES

SUPPLEMENTARY FIGURES

SUPPLEMENTARY METHODS

REFERENCES

Supplementary Tables

Table S1. Publicly available PGC GWA studies used for creation of polygenic scores

Trait	Cases	Controls	Total Sample	SNPs in common[a]	Year Published	PMID	Reference
SCZ	40675	64643	105318	487423	2018	29483656	[1]
BIP	7481	9250	16731	409796	2011	21926972	[2]
MDD[+]	59851	113154	173005	487423	2018	29700475	[3]
ASD	18381	27381	46350	492215	2017	NA	[4]
ADHD	20183	35191	55374	469352	2017	NA	[5]
OCD	2688	7037	9725	498602	2017	28761083	[6]
AN	3495	10982	14477	477354	2017	28494655	[7]
PTSD	2424	7113	9537	499383	2017	28439101	[8]

Note. [a]= number of SNPs in common between GWA summary statistics and individual-level genotypes in target sample for polygenic score creation. SCZ = Schizophrenia, BIP = Bipolar Disorder, MDD = Major Depressive Disorder, ASD = Autism Spectrum Disorder, ADHD = Attention-Deficit/Hyperactivity Disorder, OCD = Obsessive-Compulsive Disorder, AN = Anorexia Nervosa, PTSD = Post-Traumatic Stress Disorder. [+] = 23andme sample removed from discovery sample.

Table S2. Liability scale univariate SNP-heritability estimates based on LD-Score Regression

Trait	h^2	se	z-h^2	Lambda GC	Mean Chi²	Intercept	Population prev	Sample prev
SCZ	0.242	0.009	26.889	1.679	1.932	1.063	0.010	0.386
BIP	0.250	0.023	10.810	1.159	1.183	1.015	0.010	0.447
MDD	0.091	0.006	15.167	1.237	1.265	0.995	0.150	0.346
ASD	0.120	0.011	10.909	1.077	1.089	1.007	0.010	0.397
ADHD	0.223	0.016	13.936	1.270	1.319	1.016	0.050	0.364
OCD	0.243	0.038	6.460	1.05	1.056	0.993	0.015	0.276
AN	0.182	0.029	6.313	1.077	1.079	1.009	0.010	0.241
PTSD	0.132	0.059	2.253	1.017	1.013	0.994	0.080	0.254

Note. h^2 = SNP-heritability derived from LD-Score Regression. z-h^2 = z-score heritability (h^2/se). Intercept = LDSC intercept. prev = prevalences used to calculate h^2. Sample prevalence = cases/(cases+controls) based on GWAS. SCZ = Schizophrenia, BIP = Bipolar Disorder, MDD = Major Depressive Disorder, ASD = Autism Spectrum Disorder, ADHD = Attention-Deficit/Hyperactivity Disorder, OCD = Obsessive-Compulsive Disorder, AN = Anorexia Nervosa, PTSD = Post-Traumatic Stress Disorder.

Table S3. Liability scale univariate SNP-heritability estimates based on Genome-wide Complex Trait Analysis

Trait	h^2	se	$z\text{-}h^2$	Cases	Controls	Population prev	Sample prev
SCZ	0.230	0.008	28.750	9087	12171	0.01	0.43
BIP	0.250	0.012	20.833	6704	9031	0.01	0.43
MDD	0.210	0.021	10.000	9041	9381	0.15	0.49
ASD	0.170	0.025	6.800	3303	3428	0.01	0.46
ADHD	0.280	0.023	12.174	4163	12040	0.05	0.26

Note. h^2 = SNP-heritability derived from GCTA as reported in[9] . $z\text{-}h^2$ = z-score heritability (h^2/se). prev = prevalences used to calculate h^2. Sample prevalence = cases/(cases+controls) based on study sample. SCZ = Schizophrenia, BIP = Bipolar Disorder, MDD = Major Depressive Disorder, ASD = Autism Spectrum Disorder, ADHD = Attention-Deficit/Hyperactivity Disorder.

Table S4. Results from parallel analysis and initial solutions from unrotated Principal Component Analysis

		PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
Family analysis	Parallel Analysis λ[a]	**1.003**	**1.002**	1.001	1.001	1.000	1.000	0.999	0.999
	λ	**4.54**	**1.35**	0.91	0.44	0.36	0.25	0.13	0.02
	Proportion of Variance	**0.57**	**0.17**	0.11	0.06	0.05	0.03	0.02	0.00
	Cumulative Variance	**0.57**	**0.74**	0.85	0.90	0.95	0.98	1.00	1.00
GCTA	Parallel Analysis λ	-	-	-	-	-	-	-	-
	λ	**2.13**	**1.19**	0.87	0.51	0.3	-	-	-
	Proportion of Variance	**0.43**	**0.24**	0.17	0.1	0.06	-	-	-
	Cumulative Proportion	**0.43**	**0.66**	0.84	0.94	1.00	-	-	-
LDSC	Parallel Analysis λ	-	-	-	-	-	-	-	-
	λ	**2.84**	**1.85**	**1.14**	0.75	0.66	0.33	0.31	0.12
	Proportion of Variance	**0.35**	**0.23**	**0.14**	0.09	0.08	0.04	0.04	0.02
	Cumulative Proportion	**0.35**	**0.59**	**0.73**	0.82	0.90	0.95	0.98	1.00
GPS	Parallel Analysis λ[b]	**1.065**	**1.044**	1.028	1.014	1.003	0.991	0.978	0.964
	λ	**1.75**	**1.22**	1.00	1.00	0.88	0.76	0.72	0.68
	Proportion of Variance	**0.22**	**0.15**	0.13	0.12	0.11	0.10	0.09	0.09
	Cumulative Proportion	**0.22**	**0.37**	0.50	0.62	0.73	0.83	0.91	1.00

Note. λ = Eigenvalue. [a] Parameters used for parallel analysis: sample size = 3,475,112, variables = 8, repetitions = 1,000. [b] Parameters used for parallel analysis: sample size = 7,026, variables = 8, repetitions = 1,000. PC = principal component. PCs in bold represent the PCs that passed the parallel analysis selection criteria[9], or the λ > 1 criterion where parallel analysis was not possible.

Table S5. PCA solution and factor loadings for correlation between GPS based on a causal fraction of 0.10

		PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
PCA solution	Parallel Analysis λ[a]	1.065	1.044	1.028	1.014	1.003	0.991	0.978	0.964
	λ	1.71	1.19	1.00	0.99	0.88	0.82	0.74	0.68
	Proportion of Variance	0.21	0.15	0.12	0.12	0.11	0.10	0.09	0.09
	Cumulative Proportion	0.21	0.36	0.49	0.61	0.72	0.82	0.91	1.00
loadings on PC1	SCZ	**0.42***							
	BIP	**0.34***							
	MDD	**0.68***							
	ADHD	**0.57***							
	ASD	**0.55***							
	OCD	0.19*							
	AN	**0.51***							
	PTSD	0.16*							

Note. λ = Eigenvalue. *$p \leq 1.65 \times 10^{-41}$. [a]Parameters used for parallel analysis: sample size = 7,026, variables = 8, repetitions = 1,000. PC = principal component. Values printed in bold indicate factor loadings of ≥ |0.30|.

Table S6. Genome-wide association study samples used for sensitivity analyses

Trait	Cases	Controls	Total Sample	SNPs in common[a]	Year Published	PMID	Reference
SCZ	9394	12462	21856	378036	2011	21926974	[10]
BIP	7481	9250	16731	409796	2011	21926972	[2]
MDD	9240	9519	18759	374208	2013	22472876	[11]
ASD	5305	5305	10610	498435	2015	28540026	[b]
ADHD	2787	2635	5422	377332	2010	20732625	[12]
OCD	2688	7037	9725	498602	2017	28761083	[6]
AN	3495	10982	14477	477354	2017	28494655	[7]
PTSD	2424	7113	9537	499383	2017	28439101	[8]

Note. [a]= number of SNPs in common between GWA summary statistics and individual-level genotypes in target sample for polygenic score creation. [b]=interim release without publication. SCZ = Schizophrenia, BIP = Bipolar Disorder, MDD = Major Depressive Disorder, ASD = Autism Spectrum Disorder, ADHD = Attention-Deficit/Hyperactivity Disorder, OCD = Obsessive-Compulsive Disorder, AN = Anorexia Nervosa, PTSD = Post-Traumatic Stress Disorder.

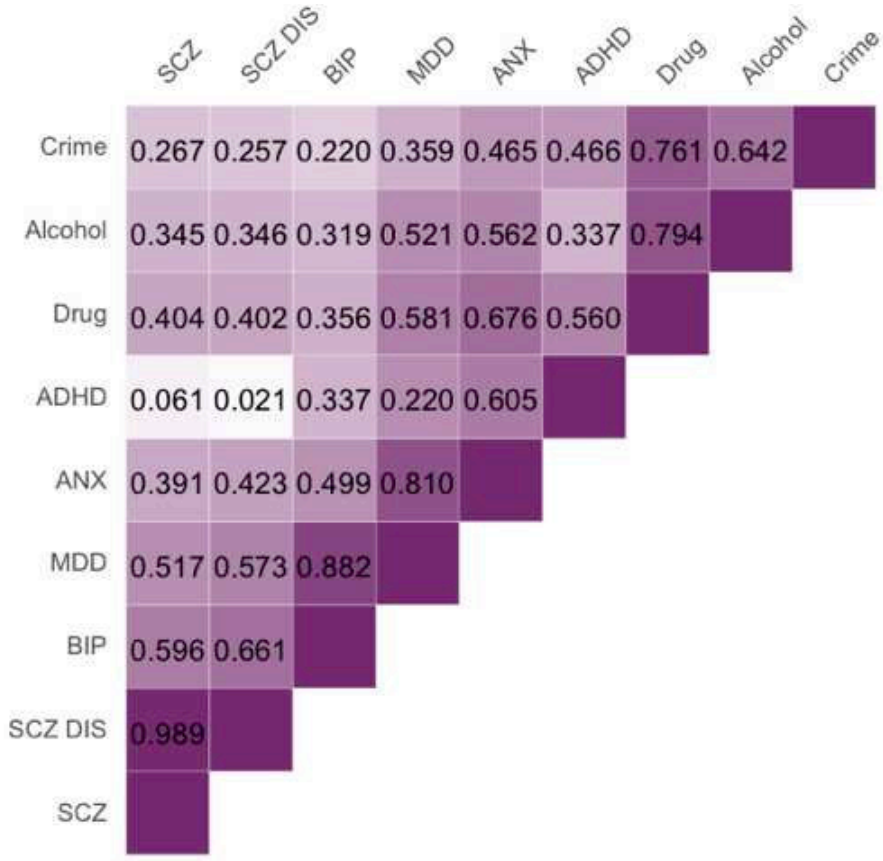
Table S7. Rotated factor loadings based on three factors for LDSC data

	RF1	RF2	RF3
SCZ	0.86	0.28	-0.02
BIP	0.84	0.14	-0.06
MDD	0.75	-0.11	0.17
ADHD	0.15	-0.39	0.77
ASD	0.52	-0.30	-0.15
OCD	0.12	0.86	0.19
AN	0.10	0.74	-0.21
PTSD	-0.05	0.25	0.91
Factor	1.00		
Correlations	0.20	1.00	
	0.18	0.00	1.00

Note. RF = rotated factor based on oblique (*Oblimin*) rotation. SCZ = Schizophrenia, BIP = Bipolar Disorder, MDD = Major Depressive Disorder, ASD = Autism Spectrum Disorder, ADHD = Attention-Deficit/Hyperactivity Disorder, OCD = Obsessive-Compulsive Disorder, AN = Anorexia Nervosa, PTSD = Post-Traumatic Stress Disorder.

Supplementary Figures

Figure S1. Original genetic correlation matrix including schizoaffective disorder as derived from family analysis



Note. SCZ = Schizophrenia, SCZ DIS = Schizoaffective Disorder, BIP = Bipolar Disorder, MDD = Major Depressive Disorder, ANX = Anxiety, ADHD = Attention-Deficit/Hyperactivity Disorder, Drug = Drug abuse, Alcohol = Alcohol abuse, Crime = Convictions of Violent Crimes.

Supplementary Methods

Methods S1. Genotyping and Quality Control

DNA for 8,122 individuals was extracted from saliva and buccal cheek swab samples and hybridized to HumanOmniExpressExome-8v1.2 genotyping arrays at the Institute of Psychiatry, Psychology and Neuroscience Genomics & Biomarker Core Facility. The raw image data from the array were normalized, pre-processed, and filtered in GenomeStudio according to Illumina Exome Chip SOP v1.4. (http://confluence.brc.iop.kcl.ac.uk:8090/display/PUB/Production+Version%3A+Illumina+Exome+Chip+SOP+v1.4). In addition, prior to genotype calling, 919 multi-mapping SNPs and 501 samples with call rate <0.95 were removed. The ZCALL program was used to augment the genotype calling for samples and SNPs that passed the initial QC.

DNA from 3,747 samples was extracted from buccal cheek swabs and genotyped at Affymetrix, Santa Clara, California, USA. From this sample, 3,665 samples were successfully hybridized to AffymetrixGeneChip 6.0 SNP genotyping arrays (http://www.affymetrix.com/support/technical/datasheets/genomewide_snp6_datasheet.pdf) using experimental protocols recommended by the manufacturer (Affymetrix Inc., Santa Clara, CA). The raw image data from the arrays were normalized and pre-processed at the Wellcome Trust Sanger Institute, Hinxton, UK for genotyping as part of the Wellcome Trust Case Control Consortium 2 (https://www.wtccc.org.uk/ccc2/) according to the manufacturer's guidelines (http://www.affymetrix.com/support/downloads/manuals/genomewidesnp6_manual.pdf). Genotypes for the Affymetrix arrays were called using CHIAMO (https://mathgen.stats.ox.ac.uk/genetics_software/chiamo/chiamo.html).

After initial quality control and genotype calling, the same quality control was performed on the samples genotyped on the Illumina and Affymetrix platforms separately using PLINK[13,14], R[15], BCFtools[16] and EIGENSOFT[17,18].

Samples were removed from subsequent analyses on the basis of call rate (<0.98), suspected non-European ancestry, heterozygosity, and relatedness other than dizygotic twin status. SNPs were excluded if the minor allele frequency was smaller than 0.5%, if more than 2% of

genotype data were missing, or if the Hardy Weinberg *p*-value was lower than 10^{-5}. Non-autosomal markers and indels were removed. Association between SNP and the platform, batch, plate or well on which samples were genotyped was calculated; SNPs with an effect *p*-value < 10^{-4} were excluded. The final sample comprised of 10,346 samples, including 7,026 unrelated individuals from which 3,320 individuals had a genotyped dizygotic co-twin. After quality control, genotype data included 4,776 individuals and 559,772 SNPs for the Illumina array, and 2,250 individuals and 635,269 SNPs for the Affymetrix array.

Genotypes from the two platforms were separately phased using EAGLE2[19], and imputed into the Haplotype Reference Consortium (release 1.1) using the Positional Burrows-Wheeler Transform method[20] through the Sanger Imputation Service[21]. Prior to merging, we excluded variants with info <0.75 and removed non-overlapping SNPs between platforms. After merging, we tested for minor allele frequency differences between platforms and removed SNPs with an effect p-value < 10^{-4}, and Hardy Weinberg p-value < 10^{-5}. Using these criteria, 7,363,646 genotyped and well-imputed SNPs were retained for the analyses. In the present study, we included unrelated individuals only (N = 7,026). To ease high computational demands by the software LDpred[22] for polygenic scoring in large samples, we further excluded SNPs with info <1, leaving 515,100 SNPs for analysis.

We performed principal component analysis on a subset of 39,353 common (MAF > 5%), perfectly imputed (info = 1) autosomal SNPs, after stringent pruning to remove markers in linkage disequilibrium (r^2 > 0.1) and excluding high linkage disequilibrium genomic regions so as to ensure that only genome-wide effects were detected.

Methods S2. Polygenic score creation using LDpred

To calculate polygenic scores, we used a Bayesian approach, *LDpred* (version 0.9.0; https://github.com/bvilhjal/ldpred/blob/master/ldpred/LDpred.py)[22]. This method has been shown to outperform predictive accuracy of the conventional clumping and *p*-value thresholding approach[22]. Using LDpred, a posterior effect size for each SNP is derived by re-weighting the original summary statistic coefficient based on (i) the relative influence of a SNP given its level of LD with surrounding SNPs, and (ii) a prior on the effect size of each SNP. This prior is dependent on the heritability of the trait, as well as the fraction of markers assumed to causally influence the trait. The final GPS is obtained as the sum of the trait-increasing alleles (each variant coded as 0,1, or 2), weighted by the posterior effect size estimates. In contrast to clumping and thresholding, LDpred retains all the SNPs in the polygenic score that are common between GWA summary statistics and genotype data in the target sample.

LDpred is computationally demanding, especially in large sample sizes with a large number of SNPs. Therefore, we restricted our analyses to 515,100 SNPs that were perfectly imputed (info score of 1). We applied a causal fraction of 1, which assumes that all SNPs contribute to the development of the trait. We decided on using this parameter to improve comparability with the other genetic methods used in this study (family analysis, GCTA, LDSC), which do not apply assumptions on the number of causally influencing SNPs and consider all genetic variations together in their estimation of genetic correlations.

Methods S3. Component selection criteria

According to the widely used Kaiser criterion[9], each PC with an eigenvalue $\lambda > 1$ represents an axis that explains more variance than a single variable itself, suggesting the retention of the component. However, randomly, uncorrelated data can produce components with $\lambda > 1$ due to chance covariation[23]. Therefore, we used parallel analysis[24] as the main criterion where possible. This method relies on the random generation of independent data, based on the same parameters as the original data (sample size; number of variables). To pass the parallel analysis criterion, eigenvalues from the study data must be larger than the 95th percentile of the distribution of the simulated components[25]. We performed PCAs to decompose four correlation matrices, one for each genetic method. Because LDSC and GCTA use thousands of SNPs to generate genetic correlations, we could only apply parallel analysis to family and GPS data, where the n number of variables used to generate a correlation matrix equates the number of variables obtained in the $n \times n$ correlation matrix. For analysis of GCTA and LDSC data, we therefore used the $\lambda > 1$ criterion instead. In addition, we performed scree plot inspection[26] to identify the point of inflection where the gradient of the line changes to a levelling-off slope, which signals that the components represent random variation rather than meaningful information.

References

1 Pardiñas AF, Holmans P, Pocklington AJ, Escott-Price V, Ripke S, Carrera N *et al.* Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat Genet* 2018; **50**: 381–389.

2 Psychiatric GWAS Consortium Bipolar Disorder Working Group. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nat Genet* 2011; **43**: 977–983.

3 Wray NR, Ripke S, Mattheisen M, Trzaskowski M, Byrne EM, Abdellaoui A *et al.* Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat Genet* 2018; **50**: 668–681.

4 Grove J, Ripke S, Als TD, Mattheisen M, Walters R, Won H *et al.* Common risk variants identified in autism spectrum disorder. *bioRxiv* 2017; : 224774.

5 Demontis D, Walters RK, Martin J, Mattheisen M, Als TD, Agerbo E *et al.* Discovery Of The First Genome-Wide Significant Risk Loci For ADHD. *bioRxiv* 2017; : 145581.

6 Arnold PD, Askland KD, Barlassina C, Bellodi L, Bienvenu OJ, Black D *et al.* Revealing the complex genetic architecture of obsessive–compulsive disorder using meta-analysis. *Molecular Psychiatry 2015 21:5* 2017; **1**: 457.

7 Duncan L, Yilmaz Z, Gaspar H, Walters R, Goldstein J, Anttila V *et al.* Significant Locus and Metabolic Genetic Correlations Revealed in Genome-Wide Association Study of Anorexia Nervosa. *American Journal of Psychiatry* 2017; **174**: 850–858.

8 Duncan LE, Ratanatharathorn A, Aiello AE, Almli LM, Amstadter AB, Ashley-Koch AE *et al.* Largest GWAS of PTSD (N=20 070) yields genetic overlap with schizophrenia and sex differences in heritability. *Molecular Psychiatry 2015 21:5* 2018; **23**: 666–673.

9 Kaiser HF. The Application of Electronic Computers to Factor Analysis. *Educational and Psychological Measurement* 1960; **20**: 141–151.

10 Ripke S, Sanders AR, Kendler KS, Levinson DF, Sklar P, Holmans PA *et al.* Genome-wide association study identifies five new schizophrenia loci. *Nat Genet* 2011; **43**: 969–976.

11 Major Depressive Disorder Working Group Consortium, Ripke S, Wray NR, Lewis CM, Hamilton SP, Weissman MM *et al.* A mega-analysis of genome-wide association studies for major depressive disorder. *Molecular Psychiatry 2015 21:5* 2013; **18**: 497–511.

12 Neale BM, Medland SE, Ripke S, Asherson P, Franke B, Lesch K-P *et al.* Meta-Analysis of Genome-Wide Association Studies of Attention-Deficit/Hyperactivity Disorder. *J Am Acad Child Adolesc Psychiatry* 2010; **49**: 884–897.

13 Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience 2015 4:1* 2015; **4**: 7.

14 Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am J Hum Genet* 2007; **81**: 559–575.

15 R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. https://www.r-project.org.

16 Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 2011; **27**: 2987–2993.

17 Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PLOS Genetics* 2006; **2**: e190.

18 Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006; **38**: 904–909.

19 Loh P-R, Danecek P, Palamara PF, Fuchsberger C, A Reshef Y, K Finucane H *et al.* Reference-based phasing using the Haplotype Reference Consortium panel. *Nat Genet* 2016; **48**: 1443–1448.

20 Durbin R. Efficient haplotype matching and storage using the positional Burrows–Wheeler transform (PBWT). *Bioinformatics* 2014; **30**: 1266–1272.

21 McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* 2016; **48**: 1279–1283.

22 Vilhjalmsson BJ, Yang J, Finucane HK, Gusev A, Lindstrom S, Ripke S *et al.* Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. *Am J Hum Genet* 2015; **97**: 576–592.

23 Jackson DA. Stopping Rules in Principal Components Analysis: A Comparison of Heuristical and Statistical Approaches. *Ecology* 1993; **74**: 2204–2214.

24 Horn JL. A rationale and test for the number of factors in factor analysis. *Psychometrika* 1965; **30**: 179–185.

25 O'connor BP. SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments, & Computers* 2000; **32**: 396–402.

26 Cattell RB. The Scree Test For The Number Of Factors. *Multivariate Behav Res* 1966; **1**: 245–276.

Appendix 4

Evidence for gene-environment correlation in child feeding: Links between common genetic variation for BMI in children and parental feeding practices

SUPPLEMENTARY METHODS

SUPPLEMENTARY TABLES

SUPPLEMENTARY FIGURES

REFERENCES

Supplementary Methods

Methods S1. Genotyping and Quality Control

DNA for 8,122 individuals (including 3,607 dizygotic co-twin samples) was extracted from saliva and buccal cheek swab samples and hybridized to HumanOmniExpressExome-8v1.2 genotyping arrays at the Institute of Psychiatry, Psychology and Neuroscience Genomics & Biomarker Core Facility. The raw image data from the array were normalized, pre-processed, and filtered in GenomeStudio according to Illumina Exome Chip SOP v1.4. (http://confluence.brc.iop.kcl.ac.uk:8090/display/PUB/Production+Version%3A+Illumina+Exome+Chip+SOP+v1.4). In addition, prior to genotype calling, 919 multi-mapping SNPs and 501 samples with call rate <0.95 were removed. The ZCALL program was used to augment the genotype calling for samples and SNPs that passed the initial QC.

DNA from 3,747 samples was extracted from buccal cheek swabs and genotyped at Affymetrix, Santa Clara, California, USA. From this sample, 3,665 samples were successfully hybridized to AffymetrixGeneChip 6.0 SNP genotyping arrays (http://www.affymetrix.com/support/technical/datasheets/genomewide_snp6_datasheet.pdf) using experimental protocols recommended by the manufacturer (Affymetrix Inc., Santa Clara, CA). The raw image data from the arrays were normalized and pre-processed at the Wellcome Trust Sanger Institute, Hinxton, UK for genotyping as part of the Wellcome Trust Case Control Consortium 2 (https://www.wtccc.org.uk/ccc2/) according to the manufacturer's guidelines (http://www.affymetrix.com/support/downloads/manuals/genomewidesnp6_manual.pdf). Genotypes for the Affymetrix arrays were called using CHIAMO (https://mathgen.stats.ox.ac.uk/genetics_software/chiamo/chiamo.html).

After initial quality control and genotype calling, the same quality control was performed on the samples genotyped on the Illumina and Affymetrix platforms separately using PLINK[1,2], R[3], BCFtools[4], and EIGENSOFT[5,6].

Samples were removed from subsequent analyses on the basis of call rate (<0.98), suspected non-European ancestry, heterozygosity, and relatedness other than dizygotic twin status.

SNPs were excluded if the minor allele frequency was smaller than 0.5%, if more than 2% of genotype data were missing, or if the Hardy Weinberg *p*-value was lower than 10^{-5}. Non-autosomal markers and indels were removed. Association between SNP and the platform, batch, plate or well on which samples were genotyped was calculated; SNPs with an effect *p*-value < 10^{-4} were excluded. A total sample of 10,346 samples (including 3,320 dizygotic twin pairs and 7,026 unrelated individuals), with 7,289 individuals and 559,772 SNPs genotyped on Illumina and 3,057 individuals and 635,269 SNPs genotyped on Affymetrix remained after quality control.

Genotypes from the two platforms were separately phased using EAGLE2[7], and imputed into the Haplotype Reference Consortium (release 1.1) using the Positional Burrows-Wheeler Transform method[8] through the Sanger Imputation Service[9]. Prior to merging, we excluded variants with info <0.75 and removed non-overlapping SNPs between platforms. After merging, we tested for minor allele frequency differences between platforms and removed SNPs with an effect p-value < 10^{-4}, and Hardy Weinberg p-value < 10^{-5}. Using these criteria, 7,363,646 genotyped and well-imputed SNPs were retained for the analyses.

We performed principal component analysis on a subset of 39,353 common (MAF > 5%), perfectly imputed (info = 1) autosomal SNPs, after stringent pruning to remove markers in linkage disequilibrium (r^2 > 0.1) and excluding high linkage disequilibrium genomic regions so as to ensure that only genome-wide effects were detected.

Supplementary Tables

Table S1a. Raw descriptive statistics of phenotypic measures in genotyped unrelated individuals

							Age effects		Sex effects					
	N	Mean	SD	Skew	Min	Max	R^2	*F*	R^2	*F*	Mean M	SD M	Mean F	SD F
Height	4442	138.85	9.54	-2.03	50.00	178.00	0.24	1378.9***	<0.01	1.51	139.03	8.58	138.68	10.35
Height SDS	4279	0.16	1.29	-3.12	-13.94	8.42	<0.01	0.02	<0.01	4.46*	0.21	1.16	0.12	1.39
Weight	4442	33.55	8.12	1.54	5.00	95.20	0.15	788.05***	<0.01	16.82***	33.03	7.29	34.03	8.78
Weight SDS	4306	0.11	1.15	-1.72	-20.13	4.41	<0.01	0.54	<0.01	2.45	0.14	1.05	0.09	1.23
BMI	4374	17.49	6.9	16.37	2.38	245.54	0.01	34.13***	<0.01	17.39***	17.04	4.14	17.91	8.91
BMI SDS	4259	-0.01	1.84	-20.34	-76.70	5.73	<0.01	0.02	<0.01	1.82	0.03	1.19	-0.05	2.28
Feeding:Pressure	4445	2.46	0.92	0.42	1.00	5.00	<0.01	18.32***	<0.01	9.94***	2.51	0.94	2.42	0.91
Feeding:Restriction	4386	3.15	1.03	-0.25	1.00	5.00	0.01	23.66***	<0.01	4.78*	3.19	1.01	3.12	1.04
Parental BMI	4112	25.91	10.61	13.11	2.58	284.80	-	-	-	-	-	-	-	-

Note. R^2 = Variance explained, *F* = *F*-statistic of ANOVA (performed on one randomly selected twin per pair). Discrepancies in sample sizes between height and weight and their respective SD scores is due to list wise deletion in the construction of SD scores due to missing age. * = $p<.05$; ** = $p<.01$; *** = $p<.001$

Table S1b. Raw descriptive statistics of phenotypic measures in genotyped DZ twins

							Age effects		Sex effects					
	N	Mean	SD	Skew	Min	Max	R^2	*F*	R^2	*F*	Mean M	SD M	Mean F	SD F
Height	4330	138.96	9.53	-2.02	52.00	178.00	0.27	757.39***	<0.01	0.96	139.15	9.80	138.75	9.36
Height SDS	4160	0.21	1.30	-3.16	-13.85	8.42	<0.01	0.51	<0.01	2.27	0.25	1.36	0.17	1.24
Weight	4310	33.67	8.33	1.53	5.00	98.00	0.17	415.3***	<0.01	3.88	33.20	7.84	33.91	8.73
Weight SDS	4186	0.15	1.23	-4.09	-28.64	4.41	<0.01	0.58	<0.01	0.29	0.15	1.41	0.12	1.11
BMI	4254	17.48	5.72	11.29	2.38	139.24	0.01	20.01***	<0.01	2.22	17.24	5.95	17.6	5.32
BMI SDS	4134	0.00	2.54	-31.40	-118.91	5.60	<0.01	0.40	<0.01	0.71	-0.09	3.96	0.02	1.28
Feeding:Pressure	4328	2.49	0.94	0.42	1.00	5.00	<0.01	6.35*	<0.01	1.21	2.50	0.94	2.45	0.94
Feeding:Restriction	4228	3.15	1.03	-0.26	1.00	5.00	<0.01	5.32*	<0.01	0.10	3.16	1.01	3.17	1.03

Note. R^2 = Variance explained, *F* = *F*-statistic of ANOVA. Discrepancies in sample sizes between height and weight and their respective SD scores is due to list wise deletion in the construction of SD scores due to missing age. Twin pairs with incomplete data were excluded. * = $p<.05$; ** = $p<.01$; *** = $p<.001$. The sample includes all DZ pairs (same sex and opposite sex); the N includes the total number of individual DZs.

Table S1c – Raw descriptive statistics of phenotypic measures in twins for twin modelling by zygosity

	N	Sample Mean	Mean m	Mean f	MZm	MZf	DZm	DZf	DZos	sex	zygosity	sex x zygosity	R^2
BMI	8630	17.50 (6.92)	17.12 (4.75)	17.86 (8.45)	17.08 (4.74)	18.15 (11.43)	17.11 (4.00)	17.70 (6.03)	17.41 (5.53)	15.61***	1.71	5.73	<0.01
BMI SDS	8406	-0.02 (2.15)	0.01 (2.23)	-0.04 (2.07)	0.03 (1.18)	-0.02 (1.27)	0.05 (1.24)	-0.05 (2.40)	-0.06 (3.10)	1.50	0.92	0.64	<0.01
Pressure	8750	2.46 (0.92)	2.50 (0.93)	2.42 (0.92)	2.43 (0.91)	2.40 (0.90)	2.56 (0.93)	2.41 (0.91)	2.49 (0.94)	5.33*	0.00	3.30	<0.01
Restriction	8582	3.15 (1.02)	3.18 (1.01)	3.13 (1.03)	3.16 (1.02)	3.10 (1.04)	3.20 (1.00)	3.13 (1.04)	3.18 (1.01)	4.17*	0.00	2.27	<0.01

Note. Means for phenotypic measures and standard deviations in brackets. MZ = monozygotic; DZ = dizygotic; m = male; f = female; os = opposite sex. F-statistics reported for sex, zygosity and sex*zygosity interaction. R^2 = variance explained by sex, zygosity and their interaction (ANOVA). Twin pairs with incomplete data and missing information about zygosity were excluded, which explains slight sample size deviations in comparison to unrelated genotyped samples. * = $p<.05$; ** = $p<.01$; *** = $p<.001$

Table S2a. Descriptive statistics of cleaned phenotypic measures (regressed onto age and sex) in unrelated individuals

							Age effects		Sex effects					
	N	Mean	SD	Skew	Min	Max	R^2	*F*	R^2	*F*	Mean M	SD M	Mean F	SD F
Height	4279	-0.01	1.00	-3.37	-11.89	5.73	<0.01	0.00	<0.01	0.00	-0.01	0.89	-0.01	1.09
Height SDS	4228	-0.02	1.00	0.09	-4.50	3.77	<0.01	0.26	<0.01	0.50	-0.03	0.98	-0.01	1.02
Weight	4306	-0.01	0.99	1.72	-4.36	7.70	<0.01	0.04	<0.01	0.04	-0.01	0.89	-0.01	1.08
Weight SDS	4228	-0.01	1.00	0.11	-4.02	3.44	<0.01	0.00	<0.01	0.03	-0.01	0.97	-0.01	1.02
BMI	4228	-0.01	0.99	2.18	-2.03	9.94	<0.01	0.10	<0.01	0.00	0.00	0.88	-0.01	1.08
BMI SDS	4228	0.00	0.99	0.14	-3.67	3.78	<0.01	0.11	<0.01	0.12	0.00	0.97	-0.01	1.01
Feeding:Pressure	4317	0.00	0.99	0.41	-1.74	2.86	<0.01	0.00	<0.01	0.12	0.00	1.01	-0.01	0.98
Feeding:Restriction	4257	0.00	1.00	-0.24	-2.26	1.95	<0.01	0.14	<0.01	0.79	0.01	0.99	-0.01	1.01
Parental BMI	3903	2.97	0.22	0.52	2.14	3.87	-	-	-	-	-	-	-	-

Note. R^2 = Variance explained, *F* = *F*-statistic of ANOVA. Discrepancies in sample sizes between height and weight and their respective SD scores is due to list wise deletion in the construction of SD scores due to missing age. * = $p<.05$; ** = $p<.01$; *** = $p<.001$

Table S2b. Descriptive statistics of cleaned phenotypic measures (regressed onto age and sex) in DZ twins

							Age effects		Sex effects					
	N	Mean	SD	Skew	Min	Max	R2	age	R2	F	Mean M	SD M	Mean F	SD F
Height	4160	0.03	1.00	-3.24	-11.25	5.73	<0.01	0.72	<0.01	0.00	0.02	1.02	0.02	0.97
Height SDS	4086	0.03	1.00	0.08	-3.05	3.76	<0.01	0.48	<0.01	1.16	0.06	0.98	0.01	1.01
Weight	4186	0.03	1.02	1.72	-4.36	7.98	<0.01	0.91	<0.01	0.09	0.03	0.96	0.01	1.06
Weight SDS	4086	0.03	1.01	0.10	-4.33	3.61	<0.01	1.17	<0.01	0.18	0.03	0.99	0.01	1.02
BMI	4086	0.02	1.01	2.04	-2.33	8.84	<0.01	0.84	<0.01	0.02	0.01	0.97	0.01	1.05
BMI SDS	4086	0.02	1.01	0.09	-4.17	3.65	<0.01	0.91	<0.01	0.06	0.00	1.02	0.01	1.03
Feeding:Pressure	4208	0.02	1.01	0.42	-1.74	2.86	<0.01	0.01	<0.01	0.36	-0.01	1.01	0.02	1.02
Feeding:Restriction	4110	0.00	1.00	-0.26	-2.26	1.93	<0.01	0.43	<0.01	1.84	-0.03	0.98	0.03	1.00

Note. R^2 = Variance explained, *F* = *F*-statistic of ANOVA (performed on one randomly selected twin per pair). Discrepancies in sample sizes between height and weight and their respective SD scores is due to list wise deletion in the construction of SD scores due to missing age. Twin pairs with incomplete data were excluded. * = $p<.05$; ** = $p<.01$; *** = $p<.001$. The sample includes all DZ pairs (same sex and opposite sex); the N includes the total number of individual DZs.

Table S2c. Descriptive statistics of cleaned phenotypic measures (regressed onto age and sex) in twins used for twin modelling by zygosity

	N	Sample Mean	Mean M	Mean F	MZm	MZf	DZm	DZf	DZos	sex	zygosity	sex x zygosity	R^2
BMI	8318	-0.00 (0.99)	-0.00 (0.91)	-0.00 (1.06)	-0.03 (0.83)	-0.04 (1.05)	0.02 (0.92)	0.03 (1.08)	0.01 (1.03)	0.48	1.24	0.98	<0.01
BMI SDS	8318	0.00 (1.00)	0.00 (0.99)	-0.00 (1.01)	-0.00 (0.95)	-0.03 (0.99)	0.02 (0.99)	0.02 (1.05)	0.00 (1.01)	0.14	0.22	0.65	<0.01
Pressure	8498	0.00 (1.00)	-0.00 (1.01)	0.00 (0.99)	-0.07 (0.99)	-0.03 (0.97)	0.08 (1.01)	-0.01 (0.99)	0.03 (1.02)	0.04	1.07	1.39	<0.01
Restriction	8332	-0.00 (1.00)	-0.00 (0.99)	0.00 (1.01)	-0.01 (1.00)	-0.03 (1.01)	0.02 (0.98)	-0.00 (1.03)	0.02 (0.99)	0.06	0.36	0.67	<0.01

Note. Means for phenotypic measures and standard deviations in brackets. MZ = monozygotic; DZ = dizygotic; m = male; f = female; os = opposite sex. F-statistics reported for sex, zygosity and sex*zygosity interaction. R^2 = variance explained by sex, zygosity and their interaction (ANOVA). Twin pairs with incomplete data and missing information about zygosity were excluded, which explains slight sample size deviations in comparison to unrelated genotyped samples. * = $p<.05$; ** = $p<.01$; *** = $p<.001$

Table S3. Twin intra-class correlations by sex and zygosity groups

	MZ males	MZ females	DZ males	DZ females	DZ opposite sex
child BMI SDS	0.84 [0.82, 0.85]	0.88 [0.87, 0.90]	0.47 [0.42, 0.51]	0.55 [0.51, 0.59]	0.47 [0.44, 0.51]
parental restriction	0.94 [0.93, 0.95]	0.94 [0.94, 0.95]	0.69 [0.66, 0.72]	0.76 [0.74, 0.78]	0.72 [0.70, 0.74]
parental pressure	0.91 [0.90, 0.92]	0.90 [0.89, 0.91]	0.60 [0.57, 0.64]	0.61 [0.57, 0.64]	0.68 [0.66, 0.70]

Note. ACE estimates are based on no sex-limitation models (for details see Supplementary Table S4).

Table S4. Fit statistics for the multivariate model including child BMI SDS, parental pressure and parental restriction

		ep	-2LL	df	AIC	χ^2	df diff	p
Saturated model		135	55259.58	24175	6909.58	-	-	-
Saturated model	Saturated model constrained	39	55350.98	24271	6808.98	91.41	96	0.61
Saturated model Constrained	ACE model with scalar	27	55401.70	24283	6835.70	50.72	12	1.04×10^{-06}
ACE model with scalar	**ACE model no scalar**	24	55403.00	24286	6831.00	1.30	3	0.73
ACE model no scalar	AE model no scalar	18	56242.27	24292	7658.27	839.26	6	5.05×10^{-178}
ACE model no scalar	CE model no scalar	18	57814.38	24292	9230.38	2411.37	6	0.00
ACE model no scalar	E model no scalar	12	67792.80	24298	19196.80	12389.79	12	0.00

Note. ep = estimated parameters; -2LL = -2 log likelihood; df = degrees of freedom; AIC = Akaike Information Criterion. An ACE model without scalar (no sex-limitation) provided best fit.

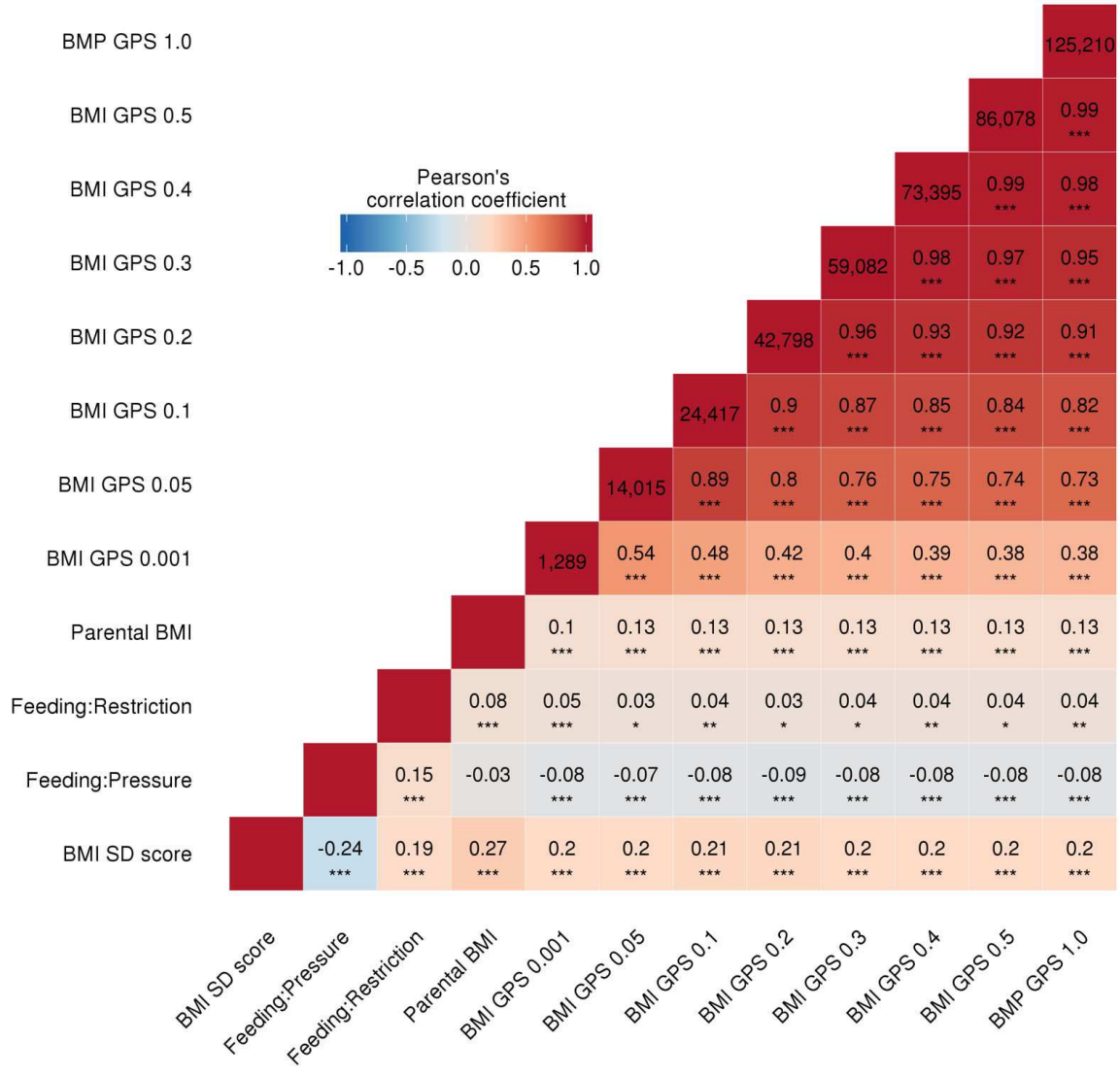
Table S5. Phenotypic, genetic, shared environmental, non-shared environmental correlations and 95% confidence intervals

	r_P [95% CI]	r_A [95% CI]	r_C [95% CI]	r_E [95% CI]
child BMI SDS & parental restriction	0.19 [0.17, 0.22]	0.28 [0.23, 0.32]	0.08 [-0.11, 0.23]	0.20 [0.15, 0.25]
child BMI SDS & parental pressure	-0.24 [-0.27, -0.22]	-0.48 [-0.52, -0.44]	0.54 [0.28, 0.98]	-0.29 [-0.34, -0.25]
parental restriction & parental pressure	0.14 [0.11, 0.16]	0.01 [-0.05, 0.06]	0.33 [0.25, 0.41]	-0.04 [-0.09, -0.01]

Note. r_P = Phenotypic correlation; r_A = genetic correlation; r_C = shared environmental correlation; r_E = non-shared environmental correlation. All estimates are based on maximum likelihood.

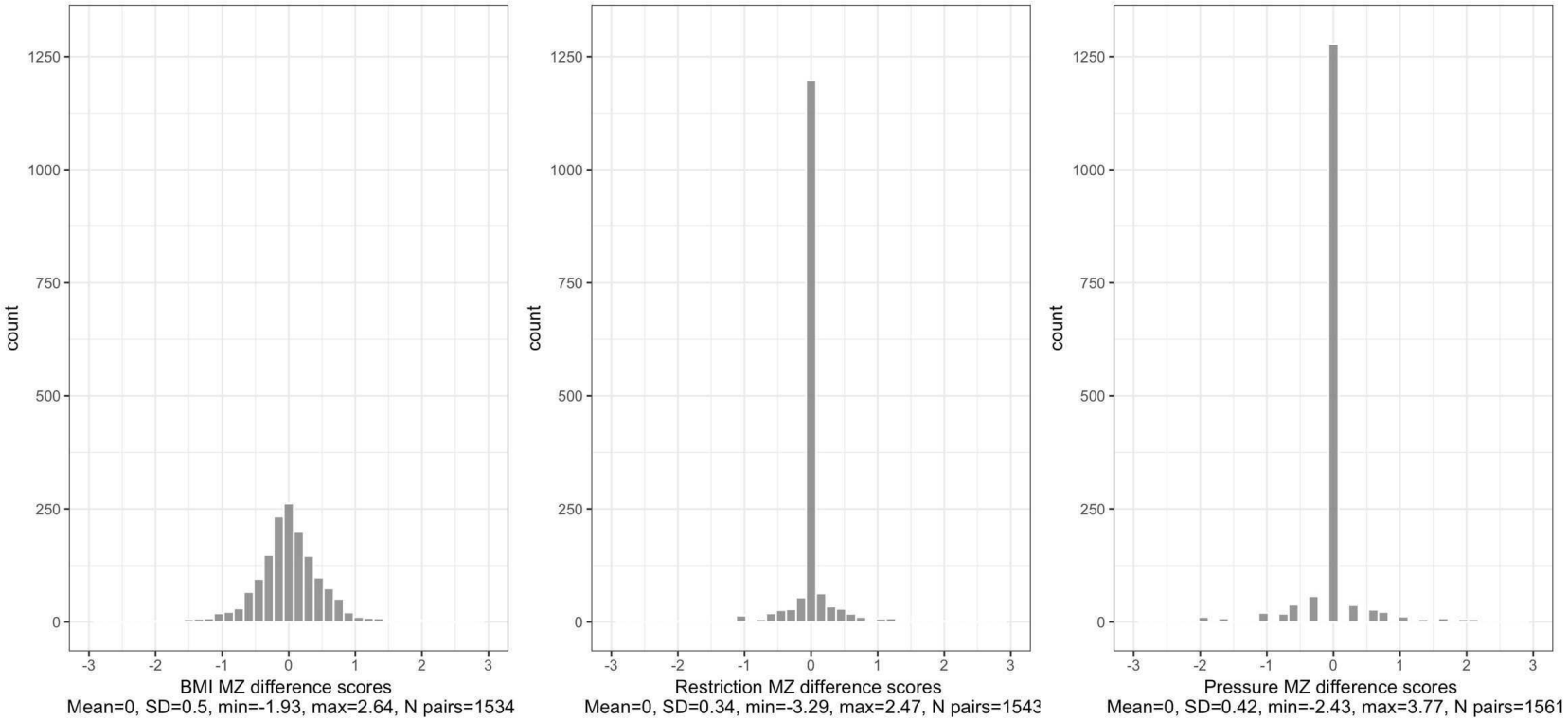
Supplementary Figures

Figure S1. Correlations across all GPS and phenotypic measures



Note. Diagonals of Genome-wide Polygenic Scores (GPS) show number of SNPs included in each respective score.

Figure S2. MZ twin difference score distributions and descriptive statistics



References

1. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience 2015 4:1. BioMed Central; 2015 Dec 1;4(1):7.

2. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. Am J Hum Genet. Cell Press; 2007 Sep 1;81(3):559–75.

3. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing [Internet]. Available from: https://www.r-project.org

4. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics. Oxford University Press; 2011 Nov 1;27(21):2987–93.

5. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. PLOS Genetics. Public Library of Science; 2006 Dec;2(12):e190.

6. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006 Aug;38(8):904–9.

7. Loh P-R, Danecek P, Palamara PF, Fuchsberger C, A Reshef Y, K Finucane H, et al. Reference-based phasing using the Haplotype Reference Consortium panel. Nat Genet. 2016 Nov;48(11):1443–8.

8. Durbin R. Efficient haplotype matching and storage using the positional Burrows–Wheeler transform (PBWT). Bioinformatics. Oxford University Press; 2014 May 1;30(9):1266–72.

9. McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, et al. A reference panel of 64,976 haplotypes for genotype imputation. Nat Genet. Nature Publishing Group; 2016 Oct;48(10):1279–83.

Appendix 5

Comparing within-and between-family polygenic score prediction

SUPPLEMENTARY METHODS

SUPPLEMENTARY TABLES

SUPPLEMENTARY FIGURES

REFERENCES

Supplementary Methods

Methods S1. Polygenic score calculation

To calculate polygenic scores, we used a Bayesian approach to polygenic score calculation, implemented in the software *LDpred*[1]. In comparison with conventional clumping and p value thresholding approaches, LDpred has demonstrated an improvement in predictive accuracy[1]. Through this method, a posterior effect size is calculated for each single SNP that is present in both the GWA study summary statistics and the target genotype sample (see Supplementary Table S2 for number of overlapping SNPs). To calculate the posterior effect size, the original summary statistic effect size estimates are adjusted based on two factors: (a) the relative influence of a SNP given its level of LD with surrounding SNPs in the target sample (here TEDS), and (b) a prior on the effect size of each SNP. To account for LD, we set the radius to a 2 megabase window. The effect size prior depends on the SNP-heritability of the discovery (i.e., GWA study) trait and an assumption on the fraction of causal markers believed to influence the discovery trait. Using the prior, the beta effect sizes are reweighted such that the effects are spread out among the SNPs across the whole genome in proportion to the LD present among these SNPs. To accommodate the high computational demands of these calculations, we reduced our genotype data set to SNPs that had perfect imputation scores (info = 1), leaving 515,100 SNPs for analysis. In the next step, all trait-associated alleles were counted (0, 1, or 2 for each SNP), weighted by the posterior SNP effect size obtained through LDpred, and summed across the genome to calculate a GPS for each individual in TEDS.

Supplementary Tables

Table S1. Descriptive statistics, age and sex effects for phenotypes

	N Pairs	Mean	SD	Skew	Min	Max	F sex	P sex	R^2 sex	F age	P age	R^2 age	P order	R^2 order
Height	1,463	171.997	10.457	0.142	132	211	1,460.639	< 0.001	0.500	0.855	0.355	0.001	0.174	<0.001
BMI	1,353	23.495	4.676	1.520	12.061	47.477	0.002	0.964	<0.001	16.470	< 0.001	0.012	0.918	<0.001
Self-rated Health	1,494	3.480	0.672	-0.330	1.000	5.000	8.188	0.004	0.005	1.457	0.228	0.001	0.987	<0.001
IQ	1,569	0.117	0.954	-0.242	-3.441	3.040	7.277	0.007	0.005	104.278	< 0.001	0.062	0.125	<0.001
GCSE	2,366	8.952	1.194	-0.330	4.670	11.000	1.814	0.178	<0.001	4.49	0.034	0.001	0.487	<0.001
Neuroticism	789	2.583	0.655	0.280	1.000	5.000	31.894	< 0.001	0.039	2.202	0.138	0.003	0.740	<0.001
ADHD Symptoms	2,469	0.063	1.002	1.357	-1.371	5.066	159.896	< 0.001	0.061	16.877	< 0.001	0.007	0.073	<0.001
SCZ Symptoms	1,140	-0.026	0.705	1.560	-0.816	4.093	2.041	0.153	0.002	6.122	0.013	0.005	0.858	<0.001
SES	2,962	0.209	0.994	0.046	-2.351	2.495	--	--	--	--	--	--	--	--

Note. Means and standard deviations for individual measures are calculated based on raw data. Height, BMI, self-reported health, GCSE grades and neuroticism means and standard deviations are reported on their original scale. IQ, ADHD symptoms, schizophrenia symptoms and socioeconomic status are reported on the z-scale as standardization was required to form the composite. Sex, age and birth order tests were performed on one randomly selected twin per pair. R^2= proportion of variance explained. Order = birth order; BMI = Body Mass Index; IQ = Intelligence; GCSE = General Certificate of Secondary Education (educational achievement); ADHD = Attention-Deficit/Hyperactivity Disorder; SCZ = Schizophrenia; SES = family socio-economic status.

Table S2. GWAS used for polygenic score calculation

Trait	Year	SNP-h^2	GPS R^2	Cases	Controls	GWAS sample size	Overlapping SNPs[a]	Notes
BMI[2]	2018	22.4% (3.7%)[2]	10.2%[2]	-	-	681,275	424,816	-
Height[2]	2018	48.3% (3.7%)[2]	24.4%[2]	-	-	693,529	427,030	-
Self-rated health[3]	2018	13% (0. 6%)[4]	--	-	-	337,199	499,422	-
Intelligence[5]	2018	19% (1%)[5]	6.7%[6]	-	-	266,453	497,059	GWAS re-run excluding TEDS sample (3,414)
Educational Attainment[7]	2018	12.2% (0.3%)[7]	11.4% [7]	-	-	766,345	496,633	-
Neuroticism[8]	2017	10.8% (0.5%)[8]	2.8% [8]	-	-	329,821	500,849	-
ADHD[9]	2019	21.6% (1.4%)[9]	5.5%[9]	20,183	35,191	55,374	469,352	-
Schizophrenia[10]	2018	20% (0.6%)[10]	5.7%[10]	40,675	64,643	105,318	487,423	-

Note. H^2 = heritability; R^2 = phenotypic variance explained; [a] = number of overlapping SNPs between GWA studies and TEDS individual-level genotype data.

Table S3. Intraclass coefficients for same-sex and opposite-sex twin pairs

	Same-sex twin pairs				Opposite-sex twin pairs			
Phenotype	**N pairs**	**ICC**	**ICC 95% CI L**	**ICC 95% CI U**	**N pairs**	**ICC**	**ICC 95% CI L**	**ICC 95% CI U**
Height	789	0.435	0.365	0.518	674	0.443	0.367	0.534
BMI	733	0.339	0.271	0.424	620	0.286	0.215	0.381
Self-rated Health	805	0.182	0.124	0.266	689	0.083	0.036	0.193
IQ	824	0.451	0.381	0.532	745	0.386	0.316	0.47
GCSE	1,220	0.579	0.517	0.647	1,146	0.585	0.521	0.656
Neuroticism	429	0.084	0.025	0.282	360	0.123	0.054	0.281
ADHD Symptoms	1,285	0.328	0.275	0.391	1,184	0.317	0.262	0.382
SCZ Symptoms	613	0.308	0.235	0.403	527	0.194	0.124	0.302

Note. BMI = Body Mass Index; IQ = Intelligence; GCSE = General Certificate of Secondary Education (educational achievement); ADHD = Attention-Deficit/Hyperactivity Disorder; SCZ = Schizophrenia symptoms; EA = Educational Attainment; Neurot = Neuroticism; ICC = Intraclass coefficient; CI = Confidence Interval.

Table S4. Intraclass coefficients for twin pairs split by genotyping chip

	Twin pairs with both twins genotyped on OEE				Twin pairs with one twin genotyped on OEE and the other twin genotyped on Affy			
Phenotype	**N pairs**	**ICC**	**ICC 95% CI L**	**ICC 95% CI U**	**N pairs**	**ICC**	**ICC 95% CI L**	**ICC 95% CI U**
Height	650	0.401	0.326	0.492	813	0.468	0.398	0.549
BMI	603	0.331	0.257	0.426	750	0.297	0.232	0.381
Self-rated Health	665	0.172	0.111	0.267	829	0.107	0.056	0.201
IQ	510	0.391	0.308	0.496	1059	0.437	0.376	0.508
GCSE	1050	0.57	0.504	0.644	1316	0.59	0.531	0.656
Neuroticism	304	0	0	0	485	0.167	0.098	0.284
ADHD Symptoms	1047	0.312	0.255	0.382	1422	0.329	0.278	0.388
SCZ Symptoms	490	0.241	0.165	0.35	650	0.263	0.195	0.356

Note. BMI = Body Mass Index; IQ = Intelligence; GCSE = General Certificate of Secondary Education (educational achievement); ADHD = Attention-Deficit/Hyperactivity Disorder; SCZ = Schizophrenia symptoms; EA = Educational Attainment; Neurot = Neuroticism; ICC = Intraclass coefficient; CI = Confidence Interval; OEE = Illumina HumanOmniExpressExome-8v1.2 chip; Affy = AffymetrixGeneChip 6.0 SNP chip.

Table S5. Intraclass coefficients

Phenotype	N pairs	ICC	ICC 95% CI L	ICC 95% CI U
Height	1,463	0.439	0.386	0.498
BMI	1,353	0.317	0.265	0.379
Self-rated Health	1,494	0.136	0.093	0.199
IQ	1,569	0.422	0.371	0.479
GCSE	2,366	0.582	0.537	0.63
Neuroticism	789	0.103	0.055	0.193
ADHD Symptoms	2,469	0.323	0.285	0.366
SCZ Symptoms	1,140	0.254	0.201	0.32

Note. BMI = Body Mass Index; IQ = Intelligence; GCSE = General Certificate of Secondary Education (educational achievement); ADHD = Attention-Deficit/Hyperactivity Disorder; SCZ = Schizophrenia symptoms; ICC = Intraclass coefficient; CI = Confidence Interval.

Table S6. Within- and between-family prediction estimates

pheno	GPS	beta.B	L.CI.B	U.CI.B	P.B	beta.W	L.CI.W	U.CI.W	P.W	TotEff	PercRed	BetaDiff	Diff.L.CI	Diff.U.CI	P.diff
ADHD	ADHD	0.111	0.082	0.14	6.80e-09	0.124	0.077	0.171	1.50e-07	0.12	-11.53	-0.013	-0.069	0.042	0.672
ADHD	BMI	0.093	0.065	0.121	6.92e-07	0.038	-0.008	0.084	0.114	0.055	59.67	0.056	0.002	0.109	0.066
ADHD	EA	-0.118	-0.145	-0.09	1.40e-10	-0.14	-0.193	-0.09	3.09e-08	-0.133	-18.66	0.022	-0.035	0.081	0.48
ADHD	Height	-0.02	-0.049	0.008	0.288	0.017	-0.028	0.061	0.464	0.005	187.39	-0.037	-0.089	0.015	0.217
ADHD	IQ	-0.106	-0.134	-0.077	1.22e-08	-0.122	-0.169	-0.074	6.38e-07	-0.117	-14.64	0.016	-0.039	0.071	0.612
ADHD	Neurot	0.044	0.015	0.073	0.021	0.052	0.008	0.098	0.023	0.05	-18.17	-0.008	-0.06	0.044	0.788
ADHD	SCZ	-0.001	-0.03	0.028	0.963	0.039	-0.007	0.083	0.107	0.026	4,500.79	-0.039	-0.093	0.013	0.195
ADHD	SRH	-0.125	-0.152	-0.097	2.78e-11	-0.035	-0.083	0.013	0.151	-0.064	71.72	-0.09	-0.145	-0.035	0.004
BMI	ADHD	0.068	0.03	0.105	0.008	0.059	-0.004	0.121	0.064	0.062	12.72	0.009	-0.064	0.082	0.833
BMI	BMI	0.35	0.313	0.385	2.35e-46	0.297	0.239	0.356	1.76e-21	0.312	15.11	0.053	-0.015	0.122	0.171
BMI	EA	-0.113	-0.152	-0.074	9.69e-06	-0.066	-0.136	0.001	0.053	-0.081	40.95	-0.046	-0.124	0.032	0.279
BMI	Height	-0.047	-0.085	-0.012	0.056	-0.032	-0.097	0.034	0.319	-0.037	32.93	-0.016	-0.091	0.058	0.699
BMI	IQ	-0.048	-0.085	-0.009	0.058	0.01	-0.058	0.077	0.76	-0.008	120.87	-0.058	-0.134	0.022	0.161
BMI	Neurot	-0.081	-0.119	-0.041	0.002	-0.007	-0.067	0.05	0.821	-0.03	91.22	-0.074	-0.144	-0.001	0.07
BMI	SCZ	-0.05	-0.091	-0.01	0.049	-0.055	-0.118	0.007	0.085	-0.054	-9.63	0.005	-0.068	0.079	0.906
BMI	SRH	-0.143	-0.183	-0.103	3.20e-08	-0.133	-0.195	-0.075	3.56e-05	-0.136	6.62	-0.009	-0.079	0.064	0.818
GCSE	ADHD	-0.176	-0.207	-0.145	7.32e-17	-0.061	-0.098	-0.025	0.001	-0.127	65.37	-0.115	-0.163	-0.068	4.95e-05
GCSE	BMI	-0.135	-0.166	-0.104	1.88e-10	-0.023	-0.061	0.016	0.239	-0.087	83.25	-0.112	-0.162	-0.063	8.10e-05
GCSE	EA	0.418	0.391	0.445	1.60e-98	0.214	0.175	0.251	2.22e-26	0.323	48.93	0.205	0.157	0.252	8.45e-14
GCSE	Height	0.033	0.002	0.065	0.108	0	-0.039	0.039	0.982	0.019	101.31	0.034	-0.016	0.083	0.233
GCSE	IQ	0.32	0.292	0.349	6.21e-55	0.201	0.163	0.239	7.24e-25	0.268	37.21	0.119	0.071	0.167	1.79e-05
GCSE	Neurot	-0.074	-0.108	-0.041	5.26e-04	-0.061	-0.097	-0.025	8.86e-04	-0.069	18.4	-0.014	-0.063	0.036	0.627
GCSE	SCZ	0.039	0.007	0.07	0.063	-0.017	-0.055	0.022	0.378	0.016	143.24	0.056	0.007	0.105	0.049
GCSE	SRH	0.188	0.157	0.219	4.93e-19	0.057	0.019	0.096	0.003	0.132	69.51	0.131	0.081	0.181	4.60e-06
SRH	ADHD	-0.056	-0.091	-0.022	0.014	-0.071	-0.137	-0.005	0.038	-0.069	-27.06	0.015	-0.059	0.091	0.713

SRH	BMI	-0.107	-0.141	-0.073	2.07e-06	-0.011	-0.076	0.054	0.752	-0.023	89.87	-0.096	-0.169	-0.024	0.019
SRH	EA	0.069	0.035	0.103	0.002	0.079	0.011	0.148	0.031	0.078	-14.17	-0.01	-0.087	0.066	0.819
SRH	Height	0.062	0.028	0.097	0.004	-0.008	-0.075	0.056	0.806	0.001	113.56	0.071	-0.003	0.143	0.082
SRH	IQ	0.022	-0.013	0.057	0.316	0.027	-0.045	0.099	0.445	0.026	-20.82	-0.005	-0.084	0.075	0.911
SRH	Neurot	-0.08	-0.114	-0.045	4.45e-04	-0.014	-0.078	0.05	0.671	-0.023	82.2	-0.066	-0.139	0.006	0.104
SRH	SCZ	-0.043	-0.078	-0.009	0.057	-0.038	-0.106	0.029	0.267	-0.039	10.71	-0.005	-0.081	0.071	0.911
SRH	SRH	0.138	0.105	0.172	1.06e-09	0.076	0.009	0.146	0.03	0.083	45.16	0.062	-0.016	0.139	0.133
Height	ADHD	-0.035	-0.076	0.006	0.176	-0.057	-0.112	-0.001	0.042	-0.047	-60.85	0.021	-0.048	0.09	0.574
Height	BMI	-0.021	-0.06	0.019	0.414	0.025	-0.028	0.079	0.368	0.005	219.22	-0.046	-0.114	0.021	0.224
Height	EA	0.034	-0.006	0.073	0.181	-0.004	-0.063	0.057	0.888	0.013	112.35	0.038	-0.032	0.109	0.33
Height	Height	0.465	0.433	0.499	7.48e-90	0.41	0.356	0.465	5.72e-53	0.432	11.83	0.055	-0.01	0.122	0.101
Height	IQ	0.042	0.002	0.081	0.098	-0.006	-0.062	0.047	0.825	0.015	115.25	0.048	-0.019	0.117	0.208
Height	Neurot	-0.058	-0.098	-0.019	0.025	0.007	-0.048	0.066	0.799	-0.022	112.01	-0.065	-0.134	0.001	0.083
Height	SCZ	-0.051	-0.09	-0.012	0.046	0.015	-0.039	0.069	0.583	-0.014	130.07	-0.067	-0.131	-0.001	0.08
Height	SRH	0.057	0.019	0.095	0.027	-0.027	-0.079	0.025	0.344	0.01	146.96	0.084	0.02	0.149	0.029
IQ	ADHD	-0.085	-0.123	-0.049	4.48e-04	-0.004	-0.057	0.049	0.874	-0.038	94.92	-0.081	-0.145	-0.016	0.027
IQ	BMI	-0.029	-0.065	0.007	0.241	0.014	-0.039	0.069	0.6	-0.004	149.71	-0.043	-0.108	0.022	0.24
IQ	EA	0.253	0.219	0.289	1.12e-26	0.126	0.068	0.184	2.15e-05	0.177	50.09	0.127	0.06	0.195	7.57e-04
IQ	Height	0.009	-0.03	0.048	0.716	0.044	-0.012	0.101	0.099	0.029	-399.06	-0.036	-0.104	0.033	0.329
IQ	IQ	0.263	0.228	0.298	2.86e-27	0.137	0.083	0.19	1.32e-06	0.187	47.98	0.126	0.061	0.192	6.27e-04
IQ	Neurot	-0.018	-0.057	0.021	0.457	-0.018	-0.072	0.037	0.496	-0.018	2	0	-0.068	0.067	0.992
IQ	SCZ	0.006	-0.031	0.043	0.824	0.012	-0.041	0.066	0.653	0.009	-122.3	-0.007	-0.071	0.058	0.854
IQ	SRH	0.109	0.071	0.145	8.19e-06	0.022	-0.033	0.08	0.435	0.058	79.42	0.086	0.019	0.151	0.021
Neurot	ADHD	-0.009	-0.057	0.04	0.765	0.081	-0.006	0.164	0.092	0.071	963.99	-0.09	-0.187	0.01	0.115
Neurot	BMI	-0.017	-0.065	0.033	0.586	0.02	-0.072	0.115	0.662	0.017	220.62	-0.037	-0.142	0.071	0.505
Neurot	EA	0.015	-0.031	0.06	0.601	0.015	-0.091	0.125	0.779	0.015	0.22	0	-0.119	0.114	1
Neurot	Height	-0.025	-0.073	0.023	0.407	0.036	-0.055	0.121	0.454	0.029	242.68	-0.06	-0.161	0.042	0.281

Neurot	IQ	-0.006	-0.05	0.038	0.849	0.057	-0.043	0.157	0.249	0.051	1,085.06	-0.063	-0.17	0.043	0.279
Neurot	Neurot	0.108	0.058	0.16	6.34e-04	0.021	-0.072	0.115	0.649	0.03	80.41	0.087	-0.016	0.191	0.122
Neurot	SCZ	-0.014	-0.064	0.038	0.662	0.063	-0.033	0.16	0.192	0.055	569.82	-0.077	-0.189	0.034	0.181
Neurot	SRH	-0.041	-0.085	0.003	0.176	-0.024	-0.121	0.075	0.617	-0.026	39.94	-0.016	-0.123	0.09	0.777
SCZ	ADHD	0.044	0.003	0.084	0.118	0.003	-0.069	0.075	0.943	0.013	94.06	0.041	-0.045	0.127	0.371
SCZ	BMI	0.109	0.069	0.149	5.72e-05	0.04	-0.036	0.115	0.276	0.057	63.06	0.068	-0.014	0.152	0.133
SCZ	EA	-0.081	-0.124	-0.038	0.003	0.009	-0.068	0.083	0.811	-0.013	111.54	-0.091	-0.176	-0.001	0.057
SCZ	Height	0	-0.043	0.041	0.991	0.002	-0.064	0.068	0.952	0.002	832.51	-0.003	-0.081	0.078	0.955
SCZ	IQ	-0.024	-0.065	0.018	0.366	0.025	-0.049	0.098	0.51	0.013	204.8	-0.049	-0.134	0.038	0.29
SCZ	Neurot	0.062	0.022	0.102	0.027	-0.024	-0.088	0.042	0.501	-0.002	138.19	0.085	0.009	0.161	0.057
SCZ	SCZ	-0.01	-0.051	0.032	0.715	-0.011	-0.085	0.064	0.777	-0.01	-4.95	0	-0.085	0.086	0.991
SCZ	SRH	-0.122	-0.164	-0.08	9.91e-06	-0.012	-0.083	0.059	0.741	-0.039	90.06	-0.11	-0.193	-0.028	0.016

Note. BMI = Body Mass Index; IQ = Intelligence; GCSE = General Certificate of Secondary Education (educational achievement); ADHD = Attention-Deficit/Hyperactivity Disorder; SCZ = Schizophrenia symptoms; EA = Educational Attainment; Neurot = Neuroticism; SRH = Self-rated Health; B = Between-family estimate; W = Within-family estimate; L.CI = Lower 95% bootstrap percentile interval; U.CI = Upper 95% bootstrap percentile interval; P = statistical significance of beta estimate; TotEff = Total effect derived as the intra-class correlation weighted sum of the within- and between family effect. PercRed = Reduction of prediction estimates when comparing within- to between-family estimates in percentage. BetaDiff = Difference of between-family and within-family beta coefficient; P.diff = statistical significance of difference between within- and between-family estimates.

Table S7. Within- and between-family prediction estimates after accounting for family socio-economic status

pheno	GPS	beta.B	L.CI.B	U.CI.B	P.B	beta.W	L.CI.W	U.CI.W	P.W	TotEff	PercRed	BetaDiff	Diff.L.CI	Diff.U.CI	P.diff
ADHD	ADHD	0.087	0.058	0.117	4.54e-06	0.122	0.076	0.17	2.50e-07	0.112	-39.79	-0.035	-0.092	0.021	0.251
ADHD	BMI	0.057	0.029	0.085	0.002	0.041	-0.006	0.088	0.088	0.046	28.76	0.016	-0.038	0.071	0.587
ADHD	EA	-0.051	-0.081	-0.021	0.01	-0.139	-0.192	-0.087	4.62e-08	-0.112	-171.34	0.088	0.028	0.15	0.006
ADHD	Height	-0.013	-0.041	0.015	0.488	0.021	-0.024	0.066	0.374	0.011	265.07	-0.034	-0.086	0.02	0.259
ADHD	IQ	-0.067	-0.096	-0.037	3.79e-04	-0.112	-0.159	-0.065	4.63e-06	-0.099	-67.12	0.045	-0.01	0.101	0.144
ADHD	Neurot	0.031	0.003	0.059	0.096	0.055	0.01	0.101	0.017	0.048	-74.95	-0.024	-0.076	0.029	0.43
ADHD	SCZ	0.002	-0.027	0.03	0.922	0.038	-0.009	0.084	0.112	0.027	-2,006.88	-0.036	-0.091	0.017	0.231
ADHD	SRH	-0.087	-0.114	-0.058	5.60e-06	-0.033	-0.08	0.016	0.183	-0.049	62.05	-0.054	-0.109	0.001	0.085
BMI	ADHD	0.041	0.004	0.078	0.107	0.054	-0.01	0.117	0.093	0.05	-32.39	-0.013	-0.086	0.06	0.746
BMI	BMI	0.328	0.292	0.364	1.10e-41	0.299	0.239	0.359	2.79e-21	0.306	9.01	0.03	-0.04	0.1	0.446
BMI	EA	-0.039	-0.082	0.003	0.15	-0.063	-0.133	0.006	0.07	-0.056	-60.15	0.024	-0.058	0.107	0.592
BMI	Height	-0.046	-0.083	-0.011	0.059	-0.03	-0.095	0.036	0.35	-0.035	34.52	-0.016	-0.091	0.058	0.694
BMI	IQ	-0.01	-0.048	0.029	0.703	0.014	-0.055	0.082	0.671	0.007	245.16	-0.024	-0.101	0.056	0.569
BMI	Neurot	-0.085	-0.123	-0.045	8.14e-04	0.004	-0.057	0.061	0.91	-0.022	104.25	-0.089	-0.16	-0.016	0.029
BMI	SCZ	-0.049	-0.088	-0.008	0.053	-0.053	-0.117	0.008	0.102	-0.052	-9.35	0.005	-0.069	0.079	0.912
BMI	SRH	-0.102	-0.142	-0.062	9.73e-05	-0.142	-0.204	-0.082	1.31e-05	-0.13	-38.44	0.039	-0.032	0.115	0.345
GCSE	ADHD	-0.083	-0.11	-0.055	5.38e-06	-0.062	-0.099	-0.025	0.001	-0.071	25.62	-0.021	-0.068	0.025	0.423
GCSE	BMI	-0.056	-0.082	-0.029	0.002	-0.024	-0.062	0.016	0.226	-0.038	57.86	-0.032	-0.079	0.016	0.224
GCSE	EA	0.243	0.216	0.27	5.23e-38	0.21	0.172	0.249	3.43e-25	0.225	13.43	0.033	-0.017	0.08	0.232
GCSE	Height	0.008	-0.019	0.035	0.639	-0.002	-0.041	0.038	0.914	0.003	125.27	0.01	-0.038	0.058	0.692
GCSE	IQ	0.212	0.187	0.238	8.65e-32	0.199	0.161	0.237	5.11e-24	0.205	5.98	0.013	-0.033	0.06	0.631
GCSE	Neurot	-0.037	-0.065	-0.009	0.043	-0.063	-0.1	-0.025	7.29e-04	-0.051	-69.83	0.026	-0.021	0.072	0.321
GCSE	SCZ	0.029	0.003	0.055	0.101	-0.018	-0.057	0.022	0.355	0.003	161.65	0.047	0	0.094	0.073
GCSE	SRH	0.069	0.041	0.098	1.87e-04	0.055	0.016	0.094	0.005	0.061	21.28	0.015	-0.034	0.065	0.584
SRH	ADHD	-0.046	-0.082	-0.011	0.046	-0.074	-0.142	-0.008	0.032	-0.071	-60.59	0.028	-0.047	0.105	0.501

SRH	BMI	-0.105	-0.14	-0.071	4.05e-06	-0.014	-0.081	0.052	0.688	-0.025	86.76	-0.091	-0.166	-0.018	0.028
SRH	EA	0.052	0.015	0.09	0.033	0.078	0.008	0.148	0.037	0.074	-48.78	-0.025	-0.106	0.053	0.567
SRH	Height	0.058	0.024	0.093	0.009	-0.007	-0.075	0.059	0.829	0.001	112.96	0.065	-0.009	0.139	0.112
SRH	IQ	0.007	-0.03	0.044	0.762	0.03	-0.042	0.103	0.396	0.027	-336.4	-0.023	-0.104	0.057	0.581
SRH	Neurot	-0.073	-0.109	-0.037	0.002	-0.016	-0.083	0.05	0.632	-0.023	77.47	-0.056	-0.132	0.019	0.171
SRH	SCZ	-0.041	-0.077	-0.006	0.071	-0.04	-0.109	0.029	0.261	-0.04	3.18	-0.001	-0.079	0.075	0.975
SRH	SRH	0.128	0.093	0.163	4.15e-08	0.075	0.006	0.147	0.034	0.081	41.88	0.054	-0.027	0.132	0.202
Height	ADHD	-0.026	-0.068	0.017	0.332	-0.061	-0.116	-0.002	0.031	-0.045	-136.37	0.035	-0.036	0.105	0.364
Height	BMI	-0.017	-0.057	0.025	0.523	0.013	-0.041	0.066	0.637	0	179.18	-0.03	-0.099	0.04	0.435
Height	EA	0.013	-0.031	0.055	0.645	-0.002	-0.062	0.059	0.96	0.005	111.66	0.014	-0.061	0.088	0.726
Height	Height	0.474	0.442	0.507	2.94e-91	0.399	0.345	0.454	1.56e-49	0.428	15.63	0.074	0.01	0.138	0.029
Height	IQ	0.027	-0.015	0.067	0.307	0.003	-0.052	0.06	0.906	0.014	87.23	0.023	-0.045	0.091	0.55
Height	Neurot	-0.055	-0.095	-0.015	0.035	0.009	-0.047	0.067	0.734	-0.019	117.14	-0.065	-0.134	0.004	0.09
Height	SCZ	-0.053	-0.093	-0.014	0.042	0.012	-0.043	0.068	0.67	-0.016	122.95	-0.065	-0.131	0.001	0.091
Height	SRH	0.047	0.008	0.086	0.082	-0.02	-0.073	0.032	0.493	0.01	141.7	0.066	0.002	0.133	0.09
IQ	ADHD	-0.041	-0.076	-0.006	0.077	-0.003	-0.057	0.053	0.923	-0.016	93.34	-0.038	-0.104	0.027	0.293
IQ	BMI	0.022	-0.012	0.055	0.346	0.009	-0.046	0.064	0.751	0.013	59.53	0.013	-0.052	0.077	0.72
IQ	EA	0.128	0.091	0.165	1.28e-07	0.119	0.058	0.177	8.18e-05	0.122	7.05	0.009	-0.06	0.079	0.815
IQ	Height	-0.006	-0.041	0.031	0.81	0.041	-0.016	0.098	0.128	0.025	850.22	-0.047	-0.113	0.02	0.187
IQ	IQ	0.191	0.158	0.227	3.28e-16	0.129	0.073	0.184	6.67e-06	0.15	32.52	0.062	-0.004	0.129	0.091
IQ	Neurot	0.002	-0.035	0.039	0.928	-0.03	-0.085	0.025	0.269	-0.019	1,526.48	0.032	-0.035	0.1	0.368
IQ	SCZ	0.008	-0.027	0.042	0.735	0.012	-0.043	0.066	0.673	0.01	-48.29	-0.004	-0.068	0.062	0.916
IQ	SRH	0.032	-0.004	0.068	0.175	0.025	-0.033	0.08	0.391	0.027	22.04	0.007	-0.057	0.074	0.85
Neurot	ADHD	-0.013	-0.062	0.038	0.686	0.083	-0.005	0.169	0.087	0.073	742.07	-0.096	-0.196	0.009	0.098
Neurot	BMI	-0.02	-0.07	0.032	0.531	0.018	-0.08	0.113	0.708	0.014	189.26	-0.038	-0.146	0.073	0.508
Neurot	EA	0.03	-0.018	0.078	0.337	0.024	-0.086	0.134	0.654	0.025	19.32	0.006	-0.114	0.129	0.926
Neurot	Height	-0.021	-0.071	0.028	0.487	0.035	-0.053	0.125	0.469	0.029	262.92	-0.056	-0.156	0.045	0.324

Neurot	IQ	-0.002	-0.049	0.045	0.959	0.074	-0.025	0.175	0.138	0.066	4,652.28	-0.076	-0.187	0.035	0.199
Neurot	Neurot	0.109	0.057	0.161	6.90e-04	0.013	-0.079	0.11	0.779	0.022	87.79	0.096	-0.013	0.202	0.094
Neurot	SCZ	-0.012	-0.065	0.039	0.689	0.053	-0.047	0.151	0.284	0.046	522.34	-0.065	-0.178	0.048	0.262
Neurot	SRH	-0.034	-0.08	0.012	0.283	-0.014	-0.115	0.085	0.775	-0.016	58.02	-0.02	-0.131	0.091	0.739
SCZ	ADHD	0.029	-0.011	0.07	0.297	0.005	-0.068	0.077	0.892	0.011	82.66	0.024	-0.062	0.112	0.604
SCZ	BMI	0.101	0.061	0.141	1.80e-04	0.043	-0.032	0.119	0.25	0.056	57.44	0.058	-0.024	0.141	0.207
SCZ	EA	-0.049	-0.097	-0.001	0.09	0.014	-0.063	0.092	0.731	-0.001	127.82	-0.063	-0.157	0.028	0.201
SCZ	Height	-0.006	-0.047	0.033	0.81	-0.002	-0.069	0.064	0.958	-0.003	68.92	-0.004	-0.084	0.075	0.924
SCZ	IQ	0.002	-0.04	0.045	0.953	0.021	-0.055	0.097	0.582	0.017	-1,242.26	-0.02	-0.108	0.071	0.676
SCZ	Neurot	0.055	0.014	0.095	0.047	-0.037	-0.1	0.027	0.305	-0.015	166.68	0.091	0.018	0.165	0.042
SCZ	SCZ	-0.004	-0.045	0.037	0.878	-0.017	-0.093	0.055	0.653	-0.014	-306.85	0.013	-0.072	0.099	0.783
SCZ	SRH	-0.1	-0.143	-0.056	3.28e-04	-0.005	-0.075	0.067	0.892	-0.027	94.96	-0.095	-0.179	-0.011	0.04

Note. BMI = Body Mass Index; IQ = Intelligence; GCSE = General Certificate of Secondary Education (educational achievement); ADHD = Attention-Deficit/Hyperactivity Disorder; SCZ = Schizophrenia symptoms; EA = Educational Attainment; Neurot = Neuroticism; SRH = Self-rated Health; B = Between-family estimate; W = Within-family estimate; L.CI = Lower 95% bootstrap percentile interval; U.CI = Upper 95% bootstrap percentile interval; P = statistical significance of beta estimate; TotEff = Total effect derived as the intra-class correlation weighted sum of the within- and between family effect. PercRed = Reduction of prediction estimates when comparing within- to between-family estimates in percentage. BetaDiff = Difference of between-family and within-family beta coefficient; P.diff = statistical significance of difference between within- and between-family estimates.

Table S8. Comparison of between-family coefficients before and after controlling for family socio-economic status

pheno	GPS	beta.B	L.CI.B	U.CI.B	P.B	beta.B.SES	L.CI.B.SES	U.CI.B.SES	P.B.SES	BetaDiff	P.Diff
ADHD	ADHD	0.111	0.082	0.14	6.80e-09	0.087	0.058	0.117	4.54e-06	0.023	0.383
ADHD	BMI	0.093	0.065	0.121	6.92e-07	0.057	0.029	0.085	0.002	0.036	0.176
ADHD	EA	-0.118	-0.145	-0.09	1.40e-10	-0.051	-0.081	-0.021	0.01	-0.067	0.013
ADHD	Height	-0.02	-0.049	0.008	0.288	-0.013	-0.041	0.015	0.488	-0.007	0.789
ADHD	IQ	-0.106	-0.134	-0.077	1.22e-08	-0.067	-0.096	-0.037	3.79e-04	-0.039	0.142
ADHD	Neurot	0.044	0.015	0.073	0.021	0.031	0.003	0.059	0.096	0.013	0.63
ADHD	SCZ	-0.001	-0.03	0.028	0.963	0.002	-0.027	0.03	0.922	-0.003	0.919
ADHD	SRH	-0.125	-0.152	-0.097	2.78e-11	-0.087	-0.114	-0.058	5.60e-06	-0.038	0.149
BMI	ADHD	0.068	0.03	0.105	0.008	0.041	0.004	0.078	0.107	0.027	0.459
BMI	BMI	0.35	0.313	0.385	2.35e-46	0.328	0.292	0.364	1.10e-41	0.022	0.515
BMI	EA	-0.113	-0.152	-0.074	9.69e-06	-0.039	-0.082	0.003	0.15	-0.073	0.049
BMI	Height	-0.047	-0.085	-0.012	0.056	-0.046	-0.083	-0.011	0.059	-0.001	0.967
BMI	IQ	-0.048	-0.085	-0.009	0.058	-0.01	-0.048	0.029	0.703	-0.038	0.285
BMI	Neurot	-0.081	-0.119	-0.041	0.002	-0.085	-0.123	-0.045	8.14e-04	0.004	0.902
BMI	SCZ	-0.05	-0.091	-0.01	0.049	-0.049	-0.088	-0.008	0.053	-0.002	0.964
BMI	SRH	-0.143	-0.183	-0.103	3.20e-08	-0.102	-0.142	-0.062	9.73e-05	-0.041	0.268
GCSE	ADHD	-0.176	-0.207	-0.145	7.32e-17	-0.083	-0.11	-0.055	5.38e-06	-0.093	7.69e-04
GCSE	BMI	-0.135	-0.166	-0.104	1.88e-10	-0.056	-0.082	-0.029	0.002	-0.079	0.004
GCSE	EA	0.418	0.391	0.445	1.60e-98	0.243	0.216	0.27	5.23e-38	0.175	3.59e-11
GCSE	Height	0.033	0.002	0.065	0.108	0.008	-0.019	0.035	0.639	0.025	0.36
GCSE	IQ	0.32	0.292	0.349	6.21e-55	0.212	0.187	0.238	8.65e-32	0.108	5.23e-05
GCSE	Neurot	-0.074	-0.108	-0.041	5.26e-04	-0.037	-0.065	-0.009	0.043	-0.037	0.183

GCSE	SCZ	0.039	0.007	0.07	0.063	0.029	0.003	0.055	0.101	0.01	0.719
GCSE	SRH	0.188	0.157	0.219	4.93e-19	0.069	0.041	0.098	1.87e-04	0.119	2.25e-05
SRH	ADHD	-0.056	-0.091	-0.022	0.014	-0.046	-0.082	-0.011	0.046	-0.01	0.766
SRH	BMI	-0.107	-0.141	-0.073	2.07e-06	-0.105	-0.14	-0.071	4.05e-06	-0.002	0.958
SRH	EA	0.069	0.035	0.103	0.002	0.052	0.015	0.09	0.033	0.017	0.603
SRH	Height	0.062	0.028	0.097	0.004	0.058	0.024	0.093	0.009	0.004	0.889
SRH	IQ	0.022	-0.013	0.057	0.316	0.007	-0.03	0.044	0.762	0.015	0.631
SRH	Neurot	-0.08	-0.114	-0.045	4.45e-04	-0.073	-0.109	-0.037	0.002	-0.007	0.826
SRH	SCZ	-0.043	-0.078	-0.009	0.057	-0.041	-0.077	-0.006	0.071	-0.002	0.95
SRH	SRH	0.138	0.105	0.172	1.06e-09	0.128	0.093	0.163	4.15e-08	0.01	0.767
Height	ADHD	-0.035	-0.076	0.006	0.176	-0.026	-0.068	0.017	0.332	-0.01	0.798
Height	BMI	-0.021	-0.06	0.019	0.414	-0.017	-0.057	0.025	0.523	-0.004	0.906
Height	EA	0.034	-0.006	0.073	0.181	0.013	-0.031	0.055	0.645	0.021	0.574
Height	Height	0.465	0.433	0.499	7.48e-90	0.474	0.442	0.507	2.94e-91	-0.008	0.792
Height	IQ	0.042	0.002	0.081	0.098	0.027	-0.015	0.067	0.307	0.015	0.679
Height	Neurot	-0.058	-0.098	-0.019	0.025	-0.055	-0.095	-0.015	0.035	-0.003	0.936
Height	SCZ	-0.051	-0.09	-0.012	0.046	-0.053	-0.093	-0.014	0.042	0.002	0.964
Height	SRH	0.057	0.019	0.095	0.027	0.047	0.008	0.086	0.082	0.01	0.782
IQ	ADHD	-0.085	-0.123	-0.049	4.48e-04	-0.041	-0.076	-0.006	0.077	-0.045	0.178
IQ	BMI	-0.029	-0.065	0.007	0.241	0.022	-0.012	0.055	0.346	-0.05	0.133
IQ	EA	0.253	0.219	0.289	1.12e-26	0.128	0.091	0.165	1.28e-07	0.126	1.69e-04
IQ	Height	0.009	-0.03	0.048	0.716	-0.006	-0.041	0.031	0.81	0.014	0.667
IQ	IQ	0.263	0.228	0.298	2.86e-27	0.191	0.158	0.227	3.28e-16	0.071	0.032
IQ	Neurot	-0.018	-0.057	0.021	0.457	0.002	-0.035	0.039	0.928	-0.02	0.545
IQ	SCZ	0.006	-0.031	0.043	0.824	0.008	-0.027	0.042	0.735	-0.002	0.945

IQ	SRH	0.109	0.071	0.145	8.19e-06	0.032	-0.004	0.068	0.175	0.077	0.023
Neurot	ADHD	-0.009	-0.057	0.04	0.765	-0.013	-0.062	0.038	0.686	0.004	0.936
Neurot	BMI	-0.017	-0.065	0.033	0.586	-0.02	-0.07	0.032	0.531	0.003	0.947
Neurot	EA	0.015	-0.031	0.06	0.601	0.03	-0.018	0.078	0.337	-0.015	0.721
Neurot	Height	-0.025	-0.073	0.023	0.407	-0.021	-0.071	0.028	0.487	-0.004	0.934
Neurot	IQ	-0.006	-0.05	0.038	0.849	-0.002	-0.049	0.045	0.959	-0.004	0.924
Neurot	Neurot	0.108	0.058	0.16	6.34e-04	0.109	0.057	0.161	6.90e-04	-0.001	0.983
Neurot	SCZ	-0.014	-0.064	0.038	0.662	-0.012	-0.065	0.039	0.689	-0.001	0.981
Neurot	SRH	-0.041	-0.085	0.003	0.176	-0.034	-0.08	0.012	0.283	-0.007	0.873
SCZ	ADHD	0.044	0.003	0.084	0.118	0.029	-0.011	0.07	0.297	0.015	0.71
SCZ	BMI	0.109	0.069	0.149	5.72e-05	0.101	0.061	0.141	1.80e-04	0.008	0.833
SCZ	EA	-0.081	-0.124	-0.038	0.003	-0.049	-0.097	-0.001	0.09	-0.032	0.418
SCZ	Height	0	-0.043	0.041	0.991	-0.006	-0.047	0.033	0.81	0.006	0.872
SCZ	IQ	-0.024	-0.065	0.018	0.366	0.002	-0.04	0.045	0.953	-0.026	0.499
SCZ	Neurot	0.062	0.022	0.102	0.027	0.055	0.014	0.095	0.047	0.007	0.861
SCZ	SCZ	-0.01	-0.051	0.032	0.715	-0.004	-0.045	0.037	0.878	-0.006	0.88
SCZ	SRH	-0.122	-0.164	-0.08	9.91e-06	-0.1	-0.143	-0.056	3.28e-04	-0.022	0.577

Note. SES = family socio-economic status; BMI = Body Mass Index; IQ = Intelligence; GCSE = General Certificate of Secondary Education (educational achievement); ADHD = Attention-Deficit/Hyperactivity Disorder; SCZ = Schizophrenia symptoms; EA = Educational Attainment; Neurot = Neuroticism; SRH = Self-rated Health; B = Between-family estimate; L.CI = Lower 95% bootstrap percentile interval; U.CI = Upper 95% bootstrap percentile interval; P = statistical significance of beta estimate; BetaDiff = Difference of the between-family beta coefficients; P.diff = statistical significance of difference between the estimates.

Table S9. Within- and between-family prediction estimates after accounting for parental education

pheno	GPS	beta.B	L.CI.B	U.CI.B	P.B	beta.W	L.CI.W	U.CI.W	P.W	TotEff	PercRed	BetaDiff	Diff.L.CI	Diff.U.CI	P.diff
ADHD	ADHD	0.091	0.059	0.122	1.00e-05	0.107	0.056	0.158	2.92e-05	0.102	-17.42	-0.016	-0.075	0.043	0.628
ADHD	BMI	0.076	0.044	0.107	1.82e-04	0.046	-0.004	0.096	0.076	0.056	39.4	0.03	-0.029	0.087	0.363
ADHD	EA	-0.048	-0.082	-0.014	0.025	-0.134	-0.19	-0.078	1.01e-06	-0.106	-177.91	0.086	0.02	0.151	0.014
ADHD	Height	-0.016	-0.047	0.015	0.428	0.015	-0.033	0.064	0.553	0.005	195.03	-0.031	-0.088	0.026	0.338
ADHD	IQ	-0.052	-0.083	-0.02	0.012	-0.111	-0.163	-0.06	3.00e-05	-0.092	-115.23	0.059	-0.001	0.121	0.076
ADHD	Neurot	0.035	0.005	0.065	0.087	0.046	-0.001	0.094	0.061	0.043	-32.09	-0.011	-0.065	0.043	0.726
ADHD	SCZ	0.02	-0.012	0.05	0.33	0.03	-0.022	0.08	0.249	0.027	-49.58	-0.01	-0.068	0.049	0.764
ADHD	SRH	-0.079	-0.109	-0.048	1.42e-04	-0.037	-0.09	0.016	0.172	-0.05	52.96	-0.042	-0.102	0.018	0.221
BMI	ADHD	0.056	0.016	0.094	0.046	0.046	-0.023	0.114	0.194	0.049	17.41	0.01	-0.069	0.088	0.83
BMI	BMI	0.353	0.313	0.393	5.37e-40	0.309	0.246	0.373	3.08e-19	0.32	12.38	0.044	-0.031	0.117	0.303
BMI	EA	-0.052	-0.098	-0.007	0.079	-0.058	-0.136	0.018	0.125	-0.057	-11.75	0.006	-0.083	0.096	0.899
BMI	Height	-0.056	-0.095	-0.017	0.038	-0.066	-0.138	0.007	0.063	-0.063	-17.4	0.01	-0.076	0.09	0.826
BMI	IQ	-0.006	-0.048	0.037	0.837	0.03	-0.045	0.106	0.405	0.02	633.96	-0.036	-0.122	0.052	0.43
BMI	Neurot	-0.086	-0.126	-0.044	0.002	-0.029	-0.097	0.036	0.399	-0.046	65.87	-0.056	-0.135	0.022	0.204
BMI	SCZ	-0.058	-0.1	-0.014	0.039	-0.073	-0.142	-0.004	0.04	-0.069	-26.64	0.015	-0.069	0.097	0.733
BMI	SRH	-0.109	-0.152	-0.066	1.53e-04	-0.101	-0.172	-0.034	0.006	-0.104	6.62	-0.007	-0.088	0.073	0.877
GCSE	ADHD	-0.115	-0.144	-0.086	9.27e-09	-0.054	-0.094	-0.013	0.011	-0.083	53.4	-0.061	-0.113	-0.011	0.034
GCSE	BMI	-0.087	-0.117	-0.058	1.37e-05	-0.012	-0.053	0.029	0.569	-0.048	86.22	-0.075	-0.126	-0.025	0.01
GCSE	EA	0.263	0.233	0.292	2.65e-37	0.213	0.171	0.255	2.98e-22	0.236	18.91	0.05	-0.002	0.101	0.093
GCSE	Height	0.013	-0.017	0.043	0.512	0.009	-0.034	0.052	0.66	0.011	28.69	0.004	-0.048	0.056	0.897
GCSE	IQ	0.205	0.177	0.234	7.43e-25	0.19	0.148	0.232	7.87e-19	0.198	7.28	0.015	-0.036	0.066	0.606
GCSE	Neurot	-0.042	-0.074	-0.011	0.035	-0.039	-0.08	0.001	0.049	-0.041	7.08	-0.003	-0.054	0.048	0.915
GCSE	SCZ	0.012	-0.017	0.042	0.533	-0.008	-0.051	0.035	0.714	0.002	161.13	0.02	-0.032	0.072	0.486
GCSE	SRH	0.1	0.07	0.129	8.00e-07	0.046	0.003	0.089	0.031	0.072	53.54	0.053	0	0.107	0.069
SRH	ADHD	-0.046	-0.085	-0.008	0.07	-0.095	-0.167	-0.022	0.012	-0.089	-108.05	0.05	-0.034	0.131	0.277

SRH	BMI	-0.091	-0.13	-0.053	3.03e-04	-0.02	-0.092	0.053	0.599	-0.029	78.07	-0.071	-0.154	0.009	0.118
SRH	EA	0.041	-0.001	0.082	0.127	0.099	0.021	0.176	0.015	0.091	-141.48	-0.058	-0.144	0.028	0.233
SRH	Height	0.044	0.007	0.082	0.07	-0.002	-0.075	0.07	0.95	0.004	105.41	0.047	-0.034	0.128	0.3
SRH	IQ	-0.001	-0.042	0.04	0.966	0.044	-0.035	0.121	0.267	0.038	4,149.06	-0.045	-0.132	0.044	0.337
SRH	Neurot	-0.084	-0.123	-0.044	9.08e-04	0.005	-0.063	0.072	0.902	-0.007	105.47	-0.088	-0.165	-0.011	0.049
SRH	SCZ	-0.024	-0.063	0.014	0.332	-0.015	-0.091	0.059	0.686	-0.017	36.51	-0.009	-0.091	0.075	0.846
SRH	SRH	0.121	0.082	0.16	2.64e-06	0.084	0.006	0.164	0.034	0.089	30.52	0.037	-0.052	0.125	0.433
Height	ADHD	-0.021	-0.067	0.027	0.48	-0.087	-0.151	-0.024	0.006	-0.059	-325.16	0.067	-0.013	0.144	0.12
Height	BMI	-0.019	-0.064	0.027	0.517	0.03	-0.031	0.09	0.34	0.01	261.55	-0.049	-0.124	0.027	0.253
Height	EA	0.004	-0.043	0.052	0.902	0.019	-0.05	0.085	0.578	0.013	-404.14	-0.015	-0.097	0.067	0.74
Height	Height	0.454	0.416	0.492	6.77e-68	0.411	0.353	0.473	1.60e-41	0.427	9.53	0.043	-0.032	0.117	0.256
Height	IQ	0.018	-0.026	0.062	0.528	-0.002	-0.066	0.062	0.96	0.007	109.09	0.02	-0.056	0.095	0.65
Height	Neurot	-0.044	-0.088	0	0.128	0.005	-0.06	0.07	0.877	-0.016	111.02	-0.049	-0.125	0.027	0.25
Height	SCZ	-0.063	-0.107	-0.018	0.03	0	-0.061	0.062	0.995	-0.026	99.71	-0.062	-0.137	0.011	0.146
Height	SRH	0.036	-0.007	0.079	0.222	-0.016	-0.078	0.046	0.639	0.006	143	0.052	-0.023	0.128	0.244
IQ	ADHD	-0.056	-0.093	-0.02	0.019	0	-0.056	0.058	0.993	-0.019	100.48	-0.056	-0.123	0.011	0.14
IQ	BMI	0.002	-0.033	0.038	0.942	0.028	-0.028	0.085	0.336	0.019	-1,506.76	-0.026	-0.093	0.042	0.486
IQ	EA	0.146	0.107	0.183	7.08e-09	0.118	0.055	0.181	2.04e-04	0.127	19.24	0.028	-0.046	0.101	0.486
IQ	Height	-0.01	-0.048	0.026	0.672	0.026	-0.034	0.085	0.364	0.013	354.85	-0.036	-0.105	0.032	0.333
IQ	IQ	0.181	0.145	0.217	1.47e-13	0.123	0.066	0.181	4.67e-05	0.143	32.07	0.058	-0.012	0.128	0.133
IQ	Neurot	0.016	-0.022	0.056	0.508	-0.015	-0.072	0.042	0.587	-0.004	196.42	0.031	-0.039	0.104	0.399
IQ	SCZ	-0.01	-0.046	0.026	0.673	0.01	-0.049	0.07	0.739	0.003	194.91	-0.02	-0.088	0.048	0.599
IQ	SRH	0.033	-0.004	0.071	0.178	-0.005	-0.066	0.055	0.874	0.008	114.9	0.038	-0.032	0.107	0.336
Neurot	ADHD	-0.003	-0.056	0.051	0.931	0.106	0.008	0.202	0.052	0.094	3,611.75	-0.109	-0.217	0.004	0.092
Neurot	BMI	-0.004	-0.057	0.052	0.915	0.048	-0.06	0.153	0.363	0.043	1,401.65	-0.052	-0.17	0.069	0.412
Neurot	EA	0.039	-0.015	0.095	0.259	0.009	-0.108	0.131	0.875	0.012	76.6	0.03	-0.103	0.161	0.657
Neurot	Height	-0.03	-0.084	0.024	0.381	0.017	-0.082	0.114	0.742	0.012	157.44	-0.047	-0.159	0.07	0.449

Neurot	IQ	0.005	-0.048	0.055	0.895	0.064	-0.046	0.178	0.243	0.058	-1,313.02	-0.06	-0.183	0.064	0.356
Neurot	Neurot	0.093	0.037	0.149	0.008	0.051	-0.053	0.153	0.324	0.055	45.2	0.042	-0.072	0.159	0.499
Neurot	SCZ	-0.009	-0.066	0.048	0.803	0.072	-0.036	0.183	0.175	0.064	929.63	-0.081	-0.206	0.046	0.202
Neurot	SRH	-0.032	-0.083	0.02	0.349	-0.058	-0.17	0.053	0.293	-0.056	-81.34	0.026	-0.098	0.15	0.688
SCZ	ADHD	0.054	0.012	0.097	0.07	-0.007	-0.086	0.072	0.856	0.008	113.33	0.061	-0.029	0.157	0.216
SCZ	BMI	0.083	0.04	0.126	0.005	0.052	-0.027	0.135	0.195	0.059	37	0.031	-0.058	0.115	0.537
SCZ	EA	-0.04	-0.087	0.009	0.198	0.017	-0.069	0.102	0.686	0.003	143.18	-0.058	-0.153	0.041	0.277
SCZ	Height	0.016	-0.031	0.063	0.581	0.008	-0.065	0.08	0.845	0.01	49.56	0.008	-0.08	0.098	0.875
SCZ	IQ	0.011	-0.036	0.056	0.716	0.042	-0.038	0.126	0.31	0.035	-297.12	-0.032	-0.13	0.063	0.534
SCZ	Neurot	0.041	-0.002	0.084	0.168	-0.036	-0.106	0.034	0.343	-0.017	187.42	0.077	-0.005	0.158	0.11
SCZ	SCZ	-0.016	-0.061	0.028	0.587	-0.026	-0.108	0.056	0.525	-0.023	-58.52	0.009	-0.085	0.104	0.85
SCZ	SRH	-0.109	-0.154	-0.062	3.16e-04	-0.005	-0.082	0.074	0.909	-0.029	95.68	-0.104	-0.195	-0.013	0.041

Note. BMI = Body Mass Index; IQ = Intelligence; GCSE = General Certificate of Secondary Education (educational achievement); ADHD = Attention-Deficit/Hyperactivity Disorder; SCZ = Schizophrenia symptoms; EA = Educational Attainment; Neurot = Neuroticism; SRH = Self-rated Health; B = Between-family estimate; W = Within-family estimate; L.CI = Lower 95% bootstrap percentile interval; U.CI = Upper 95% bootstrap percentile interval; P = statistical significance of beta estimate; TotEff = Total effect derived as the intra-class correlation weighted sum of the within- and between family effect. PercRed = Reduction of prediction estimates when comparing within- to between-family estimates in percentage. BetaDiff = Difference of between-family and within-family beta coefficient; P.diff = statistical significance of difference between within- and between-family estimates.

Table S10. Within- and between-family prediction estimates after accounting for parental occupation

pheno	GPS	beta.B	L.CI.B	U.CI.B	P.B	beta.W	L.CI.W	U.CI.W	P.W	TotEff	PercRed	BetaDiff	Diff.L.CI	Diff.U.CI	P.diff
ADHD	ADHD	0.104	0.073	0.135	2.72e-07	0.119	0.069	0.167	3.09e-06	0.114	-14.25	-0.015	-0.071	0.042	0.647
ADHD	BMI	0.08	0.051	0.11	6.32e-05	0.057	0.007	0.108	0.025	0.064	28.33	0.023	-0.035	0.08	0.484
ADHD	EA	-0.073	-0.103	-0.041	3.34e-04	-0.132	-0.188	-0.078	8.73e-07	-0.114	-82.11	0.06	-0.003	0.124	0.076
ADHD	Height	-0.01	-0.039	0.02	0.63	0.023	-0.027	0.072	0.363	0.013	340.61	-0.033	-0.09	0.026	0.311
ADHD	IQ	-0.074	-0.105	-0.043	1.66e-04	-0.124	-0.176	-0.072	2.66e-06	-0.108	-66.8	0.05	-0.011	0.11	0.131
ADHD	Neurot	0.034	0.004	0.064	0.096	0.05	0.003	0.096	0.042	0.045	-48.08	-0.016	-0.07	0.04	0.61
ADHD	SCZ	0.011	-0.02	0.041	0.586	0.023	-0.027	0.072	0.373	0.019	-106.77	-0.012	-0.069	0.045	0.718
ADHD	SRH	-0.079	-0.108	-0.05	9.22e-05	-0.051	-0.102	0.001	0.055	-0.06	35.76	-0.028	-0.088	0.03	0.395
BMI	ADHD	0.062	0.023	0.102	0.025	0.058	-0.01	0.124	0.099	0.059	6.47	0.004	-0.072	0.082	0.928
BMI	BMI	0.342	0.303	0.38	2.40e-38	0.316	0.254	0.382	9.16e-21	0.323	7.47	0.026	-0.048	0.097	0.541
BMI	EA	-0.089	-0.133	-0.047	0.002	-0.057	-0.132	0.019	0.129	-0.066	36.6	-0.033	-0.12	0.055	0.485
BMI	Height	-0.056	-0.094	-0.018	0.036	-0.08	-0.152	-0.009	0.022	-0.073	-43.1	0.024	-0.055	0.105	0.583
BMI	IQ	-0.042	-0.083	-0.001	0.119	0.025	-0.053	0.099	0.494	0.005	158.73	-0.067	-0.152	0.019	0.138
BMI	Neurot	-0.084	-0.125	-0.044	0.003	-0.029	-0.092	0.035	0.397	-0.045	65.37	-0.055	-0.131	0.02	0.214
BMI	SCZ	-0.051	-0.094	-0.007	0.066	-0.079	-0.148	-0.01	0.024	-0.07	-55.23	0.028	-0.052	0.112	0.528
BMI	SRH	-0.125	-0.167	-0.083	8.15e-06	-0.099	-0.168	-0.032	0.005	-0.107	20.69	-0.026	-0.102	0.053	0.567
GCSE	ADHD	-0.134	-0.165	-0.102	4.73e-10	-0.059	-0.1	-0.02	0.006	-0.098	55.61	-0.074	-0.125	-0.022	0.014
GCSE	BMI	-0.111	-0.143	-0.08	2.19e-07	-0.019	-0.06	0.023	0.381	-0.067	83.26	-0.093	-0.146	-0.04	0.002
GCSE	EA	0.341	0.311	0.37	1.91e-59	0.212	0.169	0.255	7.18e-22	0.275	37.7	0.128	0.075	0.182	1.58e-05
GCSE	Height	0.024	-0.008	0.056	0.263	0.008	-0.035	0.052	0.698	0.016	65.44	0.016	-0.038	0.069	0.604
GCSE	IQ	0.268	0.239	0.296	9.33e-38	0.199	0.158	0.24	6.75e-20	0.234	25.74	0.069	0.018	0.12	0.02
GCSE	Neurot	-0.045	-0.079	-0.012	0.038	-0.038	-0.08	0	0.058	-0.042	14.78	-0.007	-0.058	0.047	0.822
GCSE	SCZ	0.035	0.003	0.068	0.106	-0.006	-0.049	0.037	0.78	0.015	116.92	0.04	-0.014	0.095	0.176
GCSE	SRH	0.134	0.102	0.166	4.63e-10	0.051	0.007	0.095	0.019	0.094	62.32	0.084	0.028	0.139	0.006
SRH	ADHD	-0.038	-0.076	0.001	0.135	-0.093	-0.168	-0.016	0.016	-0.086	-147.98	0.056	-0.029	0.141	0.227

SRH	BMI	-0.082	-0.121	-0.043	0.001	-0.004	-0.076	0.068	0.907	-0.013	94.54	-0.077	-0.159	0.003	0.09
SRH	EA	0.025	-0.015	0.065	0.329	0.079	0.003	0.157	0.052	0.073	-215.64	-0.054	-0.14	0.03	0.261
SRH	Height	0.053	0.017	0.091	0.027	0.003	-0.069	0.075	0.946	0.009	95.09	0.051	-0.029	0.133	0.263
SRH	IQ	-0.01	-0.05	0.029	0.666	0.044	-0.034	0.123	0.269	0.037	519.22	-0.054	-0.142	0.034	0.242
SRH	Neurot	-0.082	-0.121	-0.042	0.001	0.001	-0.068	0.068	0.986	-0.009	100.79	-0.083	-0.16	-0.003	0.068
SRH	SCZ	-0.032	-0.07	0.006	0.199	-0.013	-0.088	0.063	0.738	-0.015	59.4	-0.019	-0.104	0.066	0.68
SRH	SRH	0.114	0.075	0.154	6.96e-06	0.073	-0.003	0.153	0.064	0.078	35.56	0.041	-0.046	0.127	0.388
Height	ADHD	-0.006	-0.05	0.039	0.838	-0.073	-0.138	-0.008	0.023	-0.045	-1,129.81	0.067	-0.011	0.145	0.12
Height	BMI	-0.02	-0.065	0.025	0.492	0.029	-0.03	0.088	0.357	0.009	246.79	-0.049	-0.122	0.025	0.252
Height	EA	0.001	-0.044	0.047	0.985	0.005	-0.061	0.072	0.875	0.003	-839.01	-0.005	-0.083	0.076	0.915
Height	Height	0.455	0.418	0.493	6.34e-70	0.4	0.339	0.46	3.40e-39	0.42	12.28	0.056	-0.019	0.129	0.14
Height	IQ	0.019	-0.024	0.061	0.492	-0.01	-0.075	0.053	0.759	0.002	152.88	0.029	-0.045	0.104	0.498
Height	Neurot	-0.034	-0.078	0.009	0.237	0.004	-0.059	0.07	0.886	-0.012	113.01	-0.039	-0.115	0.035	0.362
Height	SCZ	-0.062	-0.106	-0.017	0.028	-0.002	-0.062	0.058	0.939	-0.027	96.09	-0.06	-0.133	0.012	0.159
Height	SRH	0.034	-0.01	0.076	0.246	-0.008	-0.07	0.055	0.818	0.01	122.34	0.041	-0.035	0.115	0.345
IQ	ADHD	-0.086	-0.125	-0.047	5.25e-04	0.009	-0.048	0.069	0.761	-0.026	110.76	-0.095	-0.165	-0.027	0.015
IQ	BMI	-0.01	-0.046	0.027	0.68	0.027	-0.031	0.084	0.372	0.013	357.28	-0.037	-0.104	0.033	0.342
IQ	EA	0.208	0.17	0.245	1.90e-16	0.11	0.044	0.176	7.00e-04	0.144	46.89	0.097	0.022	0.173	0.017
IQ	Height	-0.008	-0.046	0.03	0.742	0.025	-0.036	0.086	0.396	0.013	398.74	-0.033	-0.105	0.038	0.391
IQ	IQ	0.228	0.193	0.264	7.25e-20	0.125	0.064	0.184	5.92e-05	0.161	44.92	0.102	0.033	0.176	0.01
IQ	Neurot	0.012	-0.028	0.052	0.628	-0.023	-0.081	0.035	0.424	-0.01	287.83	0.035	-0.036	0.107	0.357
IQ	SCZ	0.007	-0.029	0.044	0.782	0.011	-0.048	0.07	0.712	0.01	-56.39	-0.004	-0.074	0.065	0.919
IQ	SRH	0.043	0.006	0.082	0.09	0.01	-0.052	0.071	0.759	0.022	77.78	0.034	-0.037	0.106	0.405
Neurot	ADHD	-0.02	-0.073	0.032	0.554	0.098	0.002	0.193	0.071	0.084	581.74	-0.118	-0.229	-0.006	0.064
Neurot	BMI	-0.02	-0.075	0.035	0.556	0.051	-0.057	0.155	0.328	0.043	350.64	-0.072	-0.191	0.049	0.253
Neurot	EA	0.021	-0.031	0.075	0.534	0.028	-0.09	0.149	0.631	0.027	-34.39	-0.007	-0.137	0.122	0.915
Neurot	Height	-0.039	-0.094	0.015	0.254	0.008	-0.089	0.104	0.88	0.003	120.34	-0.047	-0.159	0.067	0.454

Neurot	IQ	0.005	-0.045	0.055	0.883	0.058	-0.053	0.168	0.289	0.053	-1,076.56	-0.054	-0.174	0.068	0.407
Neurot	Neurot	0.1	0.044	0.157	0.004	0.029	-0.076	0.137	0.576	0.036	71.01	0.071	-0.048	0.189	0.255
Neurot	SCZ	-0.037	-0.096	0.023	0.289	0.075	-0.033	0.18	0.158	0.062	303.95	-0.111	-0.236	0.011	0.077
Neurot	SRH	-0.035	-0.085	0.014	0.31	-0.03	-0.14	0.082	0.59	-0.03	14.59	-0.005	-0.129	0.116	0.938
SCZ	ADHD	0.031	-0.012	0.075	0.288	0.001	-0.075	0.082	0.971	0.008	95.23	0.03	-0.064	0.12	0.55
SCZ	BMI	0.087	0.042	0.131	0.002	0.036	-0.046	0.117	0.379	0.046	59.03	0.051	-0.037	0.141	0.3
SCZ	EA	-0.036	-0.08	0.009	0.224	0.038	-0.044	0.125	0.377	0.022	206.83	-0.074	-0.172	0.022	0.157
SCZ	Height	0.004	-0.043	0.05	0.879	-0.011	-0.083	0.063	0.791	-0.008	357.03	0.015	-0.077	0.105	0.76
SCZ	IQ	-0.002	-0.046	0.043	0.95	0.042	-0.042	0.125	0.327	0.032	2,502.22	-0.043	-0.14	0.054	0.393
SCZ	Neurot	0.021	-0.023	0.064	0.484	-0.05	-0.119	0.02	0.192	-0.035	343.21	0.071	-0.01	0.154	0.143
SCZ	SCZ	-0.026	-0.07	0.019	0.359	-0.024	-0.105	0.055	0.554	-0.024	9.03	-0.002	-0.095	0.091	0.962
SCZ	SRH	-0.092	-0.135	-0.05	0.002	0.009	-0.071	0.089	0.834	-0.013	109.33	-0.101	-0.192	-0.012	0.045

Note. BMI = Body Mass Index; IQ = Intelligence; GCSE = General Certificate of Secondary Education (educational achievement); ADHD = Attention-Deficit/Hyperactivity Disorder; SCZ = Schizophrenia symptoms; EA = Educational Attainment; Neurot = Neuroticism; SRH = Self-rated Health; B = Between-family estimate; W = Within-family estimate; L.CI = Lower 95% bootstrap percentile interval; U.CI = Upper 95% bootstrap percentile interval; P = statistical significance of beta estimate; TotEff = Total effect derived as the intra-class correlation weighted sum of the within- and between family effect. PercRed = Reduction of prediction estimates when comparing within- to between-family estimates in percentage. BetaDiff = Difference of between-family and within-family beta coefficient; P.diff = statistical significance of difference between within- and between-family estimates.

Table S11. Comparison of between-family coefficients after controlling for parental education and parental occupation

pheno	GPS	beta.B.edu	L.CI.B.edu	U.CI.B.edu	P.B.edu	beta.B.occu	L.CI.B.occu	U.CI.B.occu	P.B.occu	BetaDiff	P.Diff
ADHD	ADHD	0.091	0.059	0.122	1.00e-05	0.104	0.073	0.135	2.72e-07	-0.013	0.653
ADHD	BMI	0.076	0.044	0.107	1.82e-04	0.08	0.051	0.11	6.32e-05	-0.004	0.89
ADHD	EA	-0.048	-0.082	-0.014	0.025	-0.073	-0.103	-0.041	3.34e-04	0.025	0.404
ADHD	Height	-0.016	-0.047	0.015	0.428	-0.01	-0.039	0.02	0.63	-0.006	0.821
ADHD	IQ	-0.052	-0.083	-0.02	0.012	-0.074	-0.105	-0.043	1.66e-04	0.023	0.424
ADHD	Neurot	0.035	0.005	0.065	0.087	0.034	0.004	0.064	0.096	0.001	0.961
ADHD	SCZ	0.02	-0.012	0.05	0.33	0.011	-0.02	0.041	0.586	0.009	0.756
ADHD	SRH	-0.079	-0.109	-0.048	1.42e-04	-0.079	-0.108	-0.05	9.22e-05	0.001	0.982
BMI	ADHD	0.056	0.016	0.094	0.046	0.062	0.023	0.102	0.025	-0.006	0.87
BMI	BMI	0.353	0.313	0.393	5.37e-40	0.342	0.303	0.38	2.40e-38	0.011	0.76
BMI	EA	-0.052	-0.098	-0.007	0.079	-0.089	-0.133	-0.047	0.002	0.037	0.366
BMI	Height	-0.056	-0.095	-0.017	0.038	-0.056	-0.094	-0.018	0.036	0	0.998
BMI	IQ	-0.006	-0.048	0.037	0.837	-0.042	-0.083	-0.001	0.119	0.036	0.347
BMI	Neurot	-0.086	-0.126	-0.044	0.002	-0.084	-0.125	-0.044	0.003	-0.002	0.961
BMI	SCZ	-0.058	-0.1	-0.014	0.039	-0.051	-0.094	-0.007	0.066	-0.007	0.856
BMI	SRH	-0.109	-0.152	-0.066	1.53e-04	-0.125	-0.167	-0.083	8.15e-06	0.017	0.676
GCSE	ADHD	-0.115	-0.144	-0.086	9.27e-09	-0.134	-0.165	-0.102	4.73e-10	0.019	0.523
GCSE	BMI	-0.087	-0.117	-0.058	1.37e-05	-0.111	-0.143	-0.08	2.19e-07	0.024	0.408
GCSE	EA	0.263	0.233	0.292	2.65e-37	0.341	0.311	0.37	1.91e-59	-0.078	0.006
GCSE	Height	0.013	-0.017	0.043	0.512	0.024	-0.008	0.056	0.263	-0.011	0.708
GCSE	IQ	0.205	0.177	0.234	7.43e-25	0.268	0.239	0.296	9.33e-38	-0.062	0.028
GCSE	Neurot	-0.042	-0.074	-0.011	0.035	-0.045	-0.079	-0.012	0.038	0.003	0.93
GCSE	SCZ	0.012	-0.017	0.042	0.533	0.035	0.003	0.068	0.106	-0.022	0.448
GCSE	SRH	0.1	0.07	0.129	8.00e-07	0.134	0.102	0.166	4.63e-10	-0.034	0.242
SRH	ADHD	-0.046	-0.085	-0.008	0.07	-0.038	-0.076	0.001	0.135	-0.008	0.816

SRH	BMI	-0.091	-0.13	-0.053	3.03e-04	-0.082	-0.121	-0.043	0.001	-0.009	0.797
SRH	EA	0.041	-0.001	0.082	0.127	0.025	-0.015	0.065	0.329	0.016	0.67
SRH	Height	0.044	0.007	0.082	0.07	0.053	0.017	0.091	0.027	-0.009	0.79
SRH	IQ	-0.001	-0.042	0.04	0.966	-0.01	-0.05	0.029	0.666	0.009	0.787
SRH	Neurot	-0.084	-0.123	-0.044	9.08e-04	-0.082	-0.121	-0.042	0.001	-0.002	0.96
SRH	SCZ	-0.024	-0.063	0.014	0.332	-0.032	-0.07	0.006	0.199	0.007	0.833
SRH	SRH	0.121	0.082	0.16	2.64e-06	0.114	0.075	0.154	6.96e-06	0.007	0.843
Height	ADHD	-0.021	-0.067	0.027	0.48	-0.006	-0.05	0.039	0.838	-0.015	0.721
Height	BMI	-0.019	-0.064	0.027	0.517	-0.02	-0.065	0.025	0.492	0.001	0.98
Height	EA	0.004	-0.043	0.052	0.902	0.001	-0.044	0.047	0.985	0.003	0.94
Height	Height	0.454	0.416	0.492	6.77e-68	0.455	0.418	0.493	6.34e-70	-0.001	0.967
Height	IQ	0.018	-0.026	0.062	0.528	0.019	-0.024	0.061	0.492	-0.001	0.981
Height	Neurot	-0.044	-0.088	0	0.128	-0.034	-0.078	0.009	0.237	-0.01	0.81
Height	SCZ	-0.063	-0.107	-0.018	0.03	-0.062	-0.106	-0.017	0.028	0	0.991
Height	SRH	0.036	-0.007	0.079	0.222	0.034	-0.01	0.076	0.246	0.002	0.954
IQ	ADHD	-0.056	-0.093	-0.02	0.019	-0.086	-0.125	-0.047	5.25e-04	0.03	0.377
IQ	BMI	0.002	-0.033	0.038	0.942	-0.01	-0.046	0.027	0.68	0.012	0.728
IQ	EA	0.146	0.107	0.183	7.08e-09	0.208	0.17	0.245	1.90e-16	-0.062	0.078
IQ	Height	-0.01	-0.048	0.026	0.672	-0.008	-0.046	0.03	0.742	-0.002	0.957
IQ	IQ	0.181	0.145	0.217	1.47e-13	0.228	0.193	0.264	7.25e-20	-0.046	0.181
IQ	Neurot	0.016	-0.022	0.056	0.508	0.012	-0.028	0.052	0.628	0.004	0.916
IQ	SCZ	-0.01	-0.046	0.026	0.673	0.007	-0.029	0.044	0.782	-0.017	0.623
IQ	SRH	0.033	-0.004	0.071	0.178	0.043	0.006	0.082	0.09	-0.01	0.774
Neurot	ADHD	-0.003	-0.056	0.051	0.931	-0.02	-0.073	0.032	0.554	0.017	0.723
Neurot	BMI	-0.004	-0.057	0.052	0.915	-0.02	-0.075	0.035	0.556	0.017	0.732
Neurot	EA	0.039	-0.015	0.095	0.259	0.021	-0.031	0.075	0.534	0.018	0.704
Neurot	Height	-0.03	-0.084	0.024	0.381	-0.039	-0.094	0.015	0.254	0.009	0.852

Neurot	IQ	0.005	-0.048	0.055	0.895	0.005	-0.045	0.055	0.883	0	0.993
Neurot	Neurot	0.093	0.037	0.149	0.008	0.1	0.044	0.157	0.004	-0.007	0.89
Neurot	SCZ	-0.009	-0.066	0.048	0.803	-0.037	-0.096	0.023	0.289	0.028	0.569
Neurot	SRH	-0.032	-0.083	0.02	0.349	-0.035	-0.085	0.014	0.31	0.003	0.958
SCZ	ADHD	0.054	0.012	0.097	0.07	0.031	-0.012	0.075	0.288	0.023	0.584
SCZ	BMI	0.083	0.04	0.126	0.005	0.087	0.042	0.131	0.002	-0.004	0.918
SCZ	EA	-0.04	-0.087	0.009	0.198	-0.036	-0.08	0.009	0.224	-0.005	0.916
SCZ	Height	0.016	-0.031	0.063	0.581	0.004	-0.043	0.05	0.879	0.011	0.773
SCZ	IQ	0.011	-0.036	0.056	0.716	-0.002	-0.046	0.043	0.95	0.012	0.759
SCZ	Neurot	0.041	-0.002	0.084	0.168	0.021	-0.023	0.064	0.484	0.021	0.625
SCZ	SCZ	-0.016	-0.061	0.028	0.587	-0.026	-0.07	0.019	0.359	0.01	0.803
SCZ	SRH	-0.109	-0.154	-0.062	3.16e-04	-0.092	-0.135	-0.05	0.002	-0.016	0.697

Note. SES = family socio-economic status; BMI = Body Mass Index; IQ = Intelligence; GCSE = General Certificate of Secondary Education (educational achievement); ADHD = Attention-Deficit/Hyperactivity Disorder; SCZ = Schizophrenia symptoms; EA = Educational Attainment; Neurot = Neuroticism; SRH = Self-rated Health; B = Between-family estimate; L.CI = Lower 95% bootstrap percentile interval; U.CI = Upper 95% bootstrap percentile interval; P = statistical significance of beta estimate; BetaDiff = Difference of the between-family beta coefficients; P.diff = statistical significance of difference between the estimates.

Table S12. Within- and between-family prediction estimates of family socio-economic status

pheno	GPS	beta.B	L.CI.B	U.CI.B	P.B	beta.W	L.CI.W	U.CI.W	P.W	P.Diff
SES	ADHD	-0.168	-0.214	-0.13	1.32e-28	0	0	0	1	2.69e-08
SES	BMI	-0.19	-0.216	-0.134	3.80e-38	0	0	0	1	1.80e-10
SES	EA	0.433	0.409	0.479	2.35e-204	0	0	0	1	5.47e-49
SES	Height	0.069	0.02	0.102	6.51e-06	0	0	0	1	0.025
SES	IQ	0.231	0.213	0.291	4.28e-55	0	0	0	1	7.65e-14
SES	Neurot	-0.06	-0.129	-0.043	5.35e-05	0	0	0	1	0.049
SES	SCZ	0.04	-0.015	0.066	0.008	0	0	0	1	0.201
SES	SRH	0.302	0.238	0.318	3.86e-89	0	0	0	1	2.24e-23

Note. SES = family socio-economic status; BMI = Body Mass Index; IQ = Intelligence; GCSE = General Certificate of Secondary Education (educational achievement); ADHD = Attention-Deficit/Hyperactivity Disorder; SCZ = Schizophrenia symptoms; EA = Educational Attainment; Neurot = Neuroticism; SRH = Self-rated Health; B = Between-family estimate; W = Within-family estimate; L.CI = Lower 95% bootstrap percentile interval; U.CI = Upper 95% bootstrap percentile interval; P = statistical significance of beta estimate; P.diff = statistical significance of difference between within- and between-family estimates. Fixed effects are estimated using multiple regression analysis as parental SES is invariant within DZ twin pairs and thus mixed-effects modelling is not possible.

Table S13. Within- and between-family prediction estimates for same-sex twin pairs

pheno	GPS	beta.B	L.CI.B	U.CI.B	P.B	beta.W	L.CI.W	U.CI.W	P.W	TotEff	PercRed	BetaDiff	Diff.L.CI	Diff.U.CI	P.diff
ADHD	ADHD	0.132	0.091	0.173	6.80e-07	0.165	0.1	0.231	5.86e-07	0.154	-24.89	-0.033	-0.108	0.043	0.436
ADHD	BMI	0.088	0.05	0.128	6.93e-04	0.044	-0.016	0.105	0.165	0.059	49.89	0.044	-0.027	0.116	0.284
ADHD	EA	-0.122	-0.163	-0.08	1.90e-06	-0.209	-0.28	-0.14	9.68e-10	-0.18	-71.86	0.087	0.009	0.167	0.039
ADHD	Height	-0.067	-0.105	-0.029	0.009	-0.013	-0.079	0.051	0.707	-0.03	81.19	-0.054	-0.129	0.02	0.198
ADHD	IQ	-0.107	-0.145	-0.067	3.74e-05	-0.119	-0.184	-0.053	5.43e-04	-0.115	-11.29	0.012	-0.064	0.087	0.779
ADHD	Neurot	0.043	0.005	0.082	0.095	0.09	0.028	0.153	0.005	0.075	-109.1	-0.047	-0.119	0.026	0.254
ADHD	SCZ	0.013	-0.027	0.052	0.62	0.059	-0.005	0.122	0.074	0.044	-357.45	-0.046	-0.12	0.028	0.272
ADHD	SRH	-0.133	-0.171	-0.093	2.80e-07	-0.073	-0.139	-0.006	0.033	-0.092	45.14	-0.06	-0.133	0.014	0.161
BMI	ADHD	0.077	0.029	0.127	0.029	0.031	-0.056	0.121	0.47	0.046	59.55	0.046	-0.054	0.142	0.409
BMI	BMI	0.33	0.28	0.379	1.16e-22	0.309	0.229	0.388	2.53e-14	0.316	6.14	0.02	-0.071	0.115	0.693
BMI	EA	-0.12	-0.175	-0.067	6.78e-04	-0.073	-0.162	0.015	0.101	-0.089	39.18	-0.047	-0.152	0.059	0.406
BMI	Height	-0.093	-0.142	-0.043	0.006	-0.073	-0.165	0.018	0.096	-0.079	21.27	-0.02	-0.124	0.085	0.721
BMI	IQ	-0.065	-0.119	-0.008	0.061	0.012	-0.074	0.097	0.794	-0.014	117.69	-0.077	-0.18	0.025	0.171
BMI	Neurot	-0.056	-0.105	-0.007	0.108	-0.004	-0.086	0.076	0.92	-0.022	92.5	-0.052	-0.147	0.043	0.341
BMI	SCZ	-0.017	-0.075	0.041	0.626	-0.079	-0.162	0.003	0.064	-0.058	-363.29	0.062	-0.037	0.164	0.26
BMI	SRH	-0.172	-0.223	-0.121	8.52e-07	-0.148	-0.233	-0.066	7.60e-04	-0.156	14.1	-0.024	-0.12	0.076	0.663
GCSE	ADHD	-0.188	-0.232	-0.145	1.24e-10	-0.061	-0.115	-0.008	0.025	-0.134	67.47	-0.127	-0.197	-0.057	0.001
GCSE	BMI	-0.163	-0.207	-0.118	3.37e-08	-0.023	-0.075	0.027	0.387	-0.103	86.16	-0.141	-0.21	-0.073	3.42e-04
GCSE	EA	0.414	0.375	0.453	3.49e-49	0.222	0.168	0.277	6.33e-16	0.325	46.25	0.191	0.125	0.258	5.23e-07
GCSE	Height	0.061	0.017	0.107	0.032	0.029	-0.026	0.086	0.283	0.048	52.17	0.032	-0.04	0.104	0.418
GCSE	IQ	0.335	0.296	0.373	9.93e-32	0.204	0.148	0.26	1.96e-13	0.276	39.2	0.131	0.061	0.199	7.57e-04
GCSE	Neurot	-0.085	-0.132	-0.039	0.003	-0.112	-0.162	-0.061	1.31e-05	-0.096	-31.76	0.027	-0.04	0.097	0.484
GCSE	SCZ	-0.015	-0.059	0.029	0.599	-0.005	-0.062	0.051	0.861	-0.011	69.24	-0.01	-0.078	0.06	0.789
GCSE	SRH	0.228	0.187	0.269	2.04e-15	0.067	0.013	0.123	0.014	0.157	70.55	0.161	0.091	0.228	4.47e-05
SRH	ADHD	-0.068	-0.117	-0.019	0.033	-0.086	-0.172	-0.003	0.059	-0.083	-27.45	0.019	-0.078	0.116	0.738

SRH	BMI	-0.132	-0.175	-0.088	2.23e-05	-0.032	-0.113	0.049	0.466	-0.05	75.45	-0.1	-0.191	-0.008	0.066
SRH	EA	0.065	0.017	0.113	0.04	0.073	-0.017	0.164	0.121	0.072	-13.65	-0.009	-0.11	0.091	0.876
SRH	Height	0.063	0.016	0.11	0.039	0.011	-0.076	0.096	0.813	0.02	82.35	0.052	-0.045	0.148	0.353
SRH	IQ	0.049	0.002	0.096	0.118	0.043	-0.05	0.138	0.352	0.044	10.7	0.005	-0.1	0.11	0.926
SRH	Neurot	-0.089	-0.133	-0.043	0.004	-0.025	-0.111	0.064	0.568	-0.037	71.45	-0.063	-0.162	0.033	0.241
SRH	SCZ	-0.078	-0.128	-0.03	0.013	-0.03	-0.12	0.058	0.508	-0.039	61.62	-0.048	-0.151	0.053	0.381
SRH	SRH	0.133	0.085	0.181	1.96e-05	0.129	0.035	0.227	0.006	0.13	2.93	0.004	-0.103	0.109	0.945
Height	ADHD	-0.025	-0.08	0.029	0.474	-0.067	-0.143	0.007	0.081	-0.049	-163.44	0.042	-0.05	0.135	0.426
Height	BMI	-0.026	-0.081	0.029	0.452	-0.007	-0.08	0.066	0.845	-0.015	72.1	-0.019	-0.11	0.073	0.711
Height	EA	0.033	-0.022	0.089	0.351	-0.032	-0.112	0.047	0.421	-0.004	197.81	0.065	-0.032	0.163	0.222
Height	Height	0.435	0.39	0.479	6.39e-43	0.391	0.319	0.466	4.77e-25	0.408	10.17	0.044	-0.044	0.13	0.347
Height	IQ	0.063	0.011	0.113	0.069	-0.048	-0.121	0.021	0.221	0	176.3	0.111	0.022	0.204	0.033
Height	Neurot	-0.064	-0.119	-0.007	0.064	0.024	-0.053	0.108	0.519	-0.014	137.54	-0.088	-0.186	0.006	0.083
Height	SCZ	-0.059	-0.111	-0.007	0.093	0.039	-0.038	0.116	0.305	-0.003	166.69	-0.098	-0.188	-0.007	0.058
Height	SRH	0.054	0.001	0.106	0.12	-0.024	-0.098	0.048	0.55	0.01	143.77	0.078	-0.01	0.167	0.14
IQ	ADHD	-0.079	-0.128	-0.03	0.018	-0.03	-0.102	0.043	0.423	-0.052	62.35	-0.049	-0.137	0.036	0.323
IQ	BMI	-0.01	-0.058	0.039	0.766	0.017	-0.055	0.092	0.641	0.005	268.95	-0.027	-0.116	0.061	0.586
IQ	EA	0.27	0.222	0.319	5.60e-16	0.142	0.064	0.219	2.75e-04	0.196	47.47	0.128	0.041	0.217	0.011
IQ	Height	0.06	0.006	0.116	0.073	0.049	-0.028	0.127	0.183	0.054	18.48	0.011	-0.084	0.106	0.823
IQ	IQ	0.266	0.217	0.315	1.28e-14	0.146	0.072	0.219	1.68e-04	0.197	45.14	0.12	0.031	0.211	0.019
IQ	Neurot	-0.01	-0.061	0.043	0.773	-0.082	-0.161	-0.008	0.024	-0.049	-745.2	0.073	-0.017	0.167	0.143
IQ	SCZ	-0.017	-0.066	0.036	0.625	0.005	-0.068	0.076	0.902	-0.005	127.1	-0.021	-0.104	0.066	0.673
IQ	SRH	0.114	0.065	0.164	6.16e-04	0.025	-0.052	0.103	0.517	0.065	78.18	0.089	0.001	0.181	0.079
Neurot	ADHD	-0.025	-0.09	0.039	0.552	0.059	-0.054	0.171	0.372	0.052	331.16	-0.084	-0.215	0.043	0.282
Neurot	BMI	-0.029	-0.101	0.044	0.492	0.054	-0.068	0.18	0.386	0.047	286.23	-0.083	-0.231	0.062	0.269
Neurot	EA	0.02	-0.039	0.079	0.613	0.099	-0.052	0.258	0.177	0.092	-399.84	-0.079	-0.246	0.082	0.339
Neurot	Height	-0.012	-0.078	0.054	0.767	0.018	-0.102	0.138	0.783	0.016	247.02	-0.03	-0.167	0.106	0.695

Neurot	IQ	-0.031	-0.091	0.029	0.446	0.023	-0.113	0.162	0.749	0.018	173.55	-0.054	-0.208	0.097	0.511
Neurot	Neurot	0.126	0.062	0.191	0.002	0.045	-0.084	0.172	0.483	0.051	64.32	0.081	-0.053	0.223	0.286
Neurot	SCZ	-0.036	-0.102	0.029	0.387	0.106	-0.025	0.236	0.119	0.094	396.42	-0.141	-0.289	0.006	0.074
Neurot	SRH	-0.056	-0.115	0.004	0.161	-0.011	-0.145	0.12	0.875	-0.014	80.76	-0.045	-0.188	0.104	0.568
SCZ	ADHD	0.071	0.016	0.125	0.065	0.016	-0.079	0.111	0.737	0.033	76.97	0.054	-0.061	0.169	0.378
SCZ	BMI	0.118	0.064	0.172	0.002	0.082	-0.008	0.173	0.085	0.093	30.16	0.036	-0.066	0.137	0.559
SCZ	EA	-0.064	-0.124	-0.004	0.083	-0.015	-0.114	0.084	0.763	-0.03	76.86	-0.05	-0.167	0.065	0.423
SCZ	Height	0.022	-0.031	0.073	0.553	-0.02	-0.109	0.069	0.685	-0.007	192.28	0.041	-0.062	0.146	0.497
SCZ	IQ	-0.033	-0.089	0.023	0.371	0.014	-0.087	0.116	0.77	0	144.02	-0.047	-0.162	0.07	0.442
SCZ	Neurot	0.044	-0.013	0.101	0.249	0.02	-0.067	0.106	0.667	0.027	55.02	0.024	-0.077	0.126	0.685
SCZ	SCZ	0.029	-0.028	0.086	0.446	0.032	-0.062	0.125	0.504	0.031	-12.43	-0.004	-0.108	0.103	0.953
SCZ	SRH	-0.134	-0.194	-0.071	4.49e-04	-0.086	-0.173	0.002	0.076	-0.101	35.48	-0.047	-0.16	0.063	0.44

Note. BMI = Body Mass Index; IQ = Intelligence; GCSE = General Certificate of Secondary Education (educational achievement); ADHD = Attention-Deficit/Hyperactivity Disorder; SCZ = Schizophrenia symptoms; EA = Educational Attainment; Neurot = Neuroticism; SRH = Self-rated Health; B = Between-family estimate; W = Within-family estimate; L.CI = Lower 95% bootstrap percentile interval; U.CI = Upper 95% bootstrap percentile interval; P = statistical significance of beta estimate; TotEff = Total effect derived as the intra-class correlation weighted sum of the within- and between family effect. PercRed = Reduction of prediction estimates when comparing within- to between-family estimates in percentage. BetaDiff = Difference of between-family and within-family beta coefficient; P.diff = statistical significance of difference between within- and between-family estimates.

Table S14. Within- and between-family prediction estimates after accounting for family socio-economic status for same-sex twin pairs

pheno	GPS	beta.B	L.CI.B	U.CI.B	P.B	beta.W	L.CI.W	U.CI.W	P.W	TotEff	PercRed	BetaDiff	Diff.L.CI	Diff.U.CI	P.diff
ADHD	ADHD	0.114	0.072	0.156	1.54e-05	0.168	0.101	0.235	5.04e-07	0.152	-47.81	-0.054	-0.131	0.024	0.199
ADHD	BMI	0.057	0.017	0.096	0.029	0.042	-0.019	0.102	0.197	0.046	25.77	0.015	-0.055	0.087	0.725
ADHD	EA	-0.065	-0.11	-0.021	0.018	-0.208	-0.28	-0.137	2.07e-09	-0.163	-221.41	0.143	0.061	0.226	0.001
ADHD	Height	-0.049	-0.086	-0.01	0.056	-0.006	-0.071	0.059	0.861	-0.019	87.89	-0.043	-0.12	0.033	0.31
ADHD	IQ	-0.07	-0.11	-0.029	0.008	-0.117	-0.183	-0.051	7.95e-04	-0.103	-66.61	0.047	-0.03	0.122	0.284
ADHD	Neurot	0.038	0.001	0.076	0.136	0.096	0.032	0.159	0.004	0.078	-151.34	-0.058	-0.13	0.015	0.168
ADHD	SCZ	0.014	-0.026	0.054	0.576	0.055	-0.011	0.121	0.098	0.043	-288.78	-0.041	-0.117	0.033	0.327
ADHD	SRH	-0.1	-0.138	-0.061	1.39e-04	-0.069	-0.135	-0.001	0.048	-0.078	31.15	-0.031	-0.106	0.044	0.474
BMI	ADHD	0.043	-0.003	0.089	0.215	0.023	-0.066	0.111	0.592	0.029	46.08	0.02	-0.079	0.118	0.72
BMI	BMI	0.309	0.258	0.358	1.59e-20	0.313	0.233	0.392	1.44e-14	0.312	-1.35	-0.004	-0.097	0.091	0.935
BMI	EA	-0.05	-0.11	0.007	0.179	-0.063	-0.151	0.027	0.161	-0.059	-24.56	0.012	-0.099	0.121	0.832
BMI	Height	-0.073	-0.123	-0.025	0.028	-0.068	-0.161	0.025	0.124	-0.069	7.55	-0.006	-0.112	0.101	0.92
BMI	IQ	-0.027	-0.083	0.028	0.438	0.015	-0.071	0.103	0.733	0.002	155.64	-0.042	-0.146	0.063	0.453
BMI	Neurot	-0.067	-0.117	-0.018	0.049	-0.005	-0.089	0.076	0.907	-0.024	92.56	-0.062	-0.159	0.036	0.254
BMI	SCZ	-0.023	-0.079	0.035	0.503	-0.078	-0.162	0.005	0.069	-0.061	-240.46	0.055	-0.048	0.159	0.314
BMI	SRH	-0.127	-0.177	-0.076	3.34e-04	-0.156	-0.242	-0.074	3.72e-04	-0.147	-23.13	0.029	-0.069	0.133	0.6
GCSE	ADHD	-0.097	-0.134	-0.059	1.27e-04	-0.072	-0.128	-0.018	0.01	-0.083	25.66	-0.025	-0.092	0.041	0.508
GCSE	BMI	-0.089	-0.127	-0.051	4.88e-04	-0.021	-0.073	0.03	0.425	-0.051	76.08	-0.068	-0.132	-0.002	0.066
GCSE	EA	0.247	0.209	0.284	7.33e-21	0.222	0.167	0.278	2.40e-15	0.233	10.32	0.026	-0.046	0.094	0.5
GCSE	Height	0.015	-0.021	0.052	0.54	0.03	-0.026	0.085	0.288	0.023	-94.32	-0.014	-0.08	0.052	0.7
GCSE	IQ	0.224	0.189	0.26	1.08e-18	0.207	0.152	0.264	2.34e-13	0.215	7.45	0.017	-0.051	0.084	0.656
GCSE	Neurot	-0.063	-0.1	-0.024	0.011	-0.115	-0.166	-0.064	1.52e-05	-0.091	-82.83	0.052	-0.011	0.116	0.15
GCSE	SCZ	-0.005	-0.043	0.032	0.84	-0.002	-0.059	0.053	0.927	-0.004	50.01	-0.002	-0.068	0.064	0.946
GCSE	SRH	0.11	0.073	0.145	1.70e-05	0.067	0.011	0.123	0.017	0.086	39.2	0.043	-0.025	0.109	0.254

SRH	ADHD	-0.051	-0.101	-0.003	0.114	-0.085	-0.173	-0.001	0.066	-0.079	-66.99	0.034	-0.061	0.135	0.544
SRH	BMI	-0.125	-0.169	-0.081	8.32e-05	-0.036	-0.118	0.048	0.426	-0.05	71.25	-0.089	-0.182	0.004	0.105
SRH	EA	0.038	-0.016	0.09	0.265	0.078	-0.012	0.167	0.104	0.071	-105.54	-0.04	-0.145	0.064	0.494
SRH	Height	0.052	0.006	0.1	0.086	0.014	-0.074	0.101	0.76	0.021	72.33	0.038	-0.059	0.136	0.501
SRH	IQ	0.028	-0.023	0.078	0.38	0.05	-0.041	0.144	0.287	0.047	-77.88	-0.022	-0.13	0.084	0.699
SRH	Neurot	-0.086	-0.133	-0.039	0.006	-0.027	-0.115	0.061	0.55	-0.037	68.28	-0.059	-0.159	0.043	0.288
SRH	SCZ	-0.067	-0.117	-0.017	0.033	-0.036	-0.127	0.054	0.439	-0.041	46.96	-0.032	-0.134	0.073	0.571
SRH	SRH	0.122	0.071	0.173	1.53e-04	0.131	0.031	0.227	0.006	0.129	-6.97	-0.009	-0.119	0.101	0.882
Height	ADHD	-0.013	-0.07	0.044	0.717	-0.067	-0.146	0.011	0.084	-0.044	-418.39	0.054	-0.042	0.15	0.304
Height	BMI	-0.015	-0.071	0.04	0.678	-0.013	-0.085	0.061	0.729	-0.014	11.12	-0.002	-0.095	0.089	0.975
Height	EA	-0.004	-0.064	0.056	0.92	-0.038	-0.116	0.044	0.345	-0.023	-897.21	0.034	-0.071	0.137	0.536
Height	Height	0.434	0.39	0.478	4.00e-42	0.386	0.311	0.462	8.27e-24	0.405	10.97	0.048	-0.041	0.137	0.318
Height	IQ	0.04	-0.013	0.094	0.266	-0.05	-0.122	0.024	0.212	-0.011	224.5	0.09	-0.004	0.183	0.094
Height	Neurot	-0.064	-0.12	-0.007	0.066	0.033	-0.047	0.118	0.382	-0.008	152.19	-0.098	-0.197	-0.002	0.059
Height	SCZ	-0.056	-0.11	-0.003	0.11	0.036	-0.042	0.113	0.359	-0.004	163.31	-0.092	-0.184	0	0.079
Height	SRH	0.039	-0.018	0.092	0.285	-0.023	-0.095	0.05	0.574	0.004	158.37	0.061	-0.032	0.153	0.256
IQ	ADHD	-0.04	-0.086	0.005	0.201	-0.031	-0.104	0.042	0.415	-0.034	22.33	-0.009	-0.094	0.08	0.856
IQ	BMI	0.025	-0.02	0.07	0.433	0.019	-0.056	0.096	0.607	0.021	22.72	0.006	-0.083	0.092	0.908
IQ	EA	0.157	0.105	0.208	2.97e-06	0.132	0.053	0.21	8.53e-04	0.141	15.55	0.024	-0.069	0.117	0.637
IQ	Height	0.019	-0.031	0.07	0.559	0.041	-0.038	0.12	0.273	0.033	-121.64	-0.023	-0.115	0.072	0.646
IQ	IQ	0.199	0.153	0.247	2.21e-09	0.137	0.062	0.212	5.70e-04	0.16	31.39	0.063	-0.027	0.153	0.223
IQ	Neurot	0.002	-0.046	0.051	0.954	-0.084	-0.162	-0.006	0.025	-0.052	4,776	0.086	-0.006	0.178	0.079
IQ	SCZ	-0.006	-0.052	0.04	0.852	0.008	-0.064	0.08	0.836	0.003	231.51	-0.014	-0.099	0.071	0.78
IQ	SRH	0.046	-0.003	0.095	0.146	0.021	-0.056	0.101	0.595	0.03	55.1	0.026	-0.067	0.116	0.612
Neurot	ADHD	-0.029	-0.095	0.037	0.499	0.063	-0.052	0.178	0.348	0.055	312.8	-0.092	-0.224	0.041	0.246
Neurot	BMI	-0.033	-0.107	0.045	0.455	0.05	-0.077	0.176	0.429	0.043	254.86	-0.083	-0.232	0.065	0.281
Neurot	EA	0.034	-0.032	0.098	0.429	0.096	-0.059	0.249	0.196	0.091	-181.03	-0.062	-0.227	0.104	0.47

Neurot	Height	-0.009	-0.077	0.058	0.826	0.016	-0.106	0.134	0.809	0.014	273.61	-0.026	-0.163	0.115	0.746
Neurot	IQ	-0.03	-0.095	0.036	0.481	0.03	-0.108	0.167	0.678	0.025	198.5	-0.06	-0.213	0.094	0.472
Neurot	Neurot	0.13	0.064	0.197	0.002	0.041	-0.093	0.172	0.536	0.048	68.54	0.089	-0.051	0.236	0.253
Neurot	SCZ	-0.041	-0.108	0.028	0.334	0.09	-0.04	0.217	0.194	0.079	321.75	-0.131	-0.278	0.02	0.105
Neurot	SRH	-0.056	-0.117	0.006	0.181	-0.01	-0.144	0.119	0.89	-0.013	82.75	-0.046	-0.189	0.104	0.568
SCZ	ADHD	0.058	0.004	0.115	0.125	0.018	-0.079	0.117	0.718	0.029	69.51	0.041	-0.076	0.158	0.514
SCZ	BMI	0.103	0.049	0.156	0.006	0.091	-0.001	0.182	0.06	0.095	12.02	0.012	-0.09	0.114	0.839
SCZ	EA	-0.044	-0.108	0.021	0.268	-0.018	-0.115	0.082	0.72	-0.025	58.82	-0.026	-0.145	0.089	0.687
SCZ	Height	0.034	-0.019	0.088	0.343	-0.021	-0.113	0.069	0.667	-0.006	162.63	0.056	-0.048	0.16	0.364
SCZ	IQ	-0.015	-0.07	0.041	0.689	0.011	-0.091	0.113	0.826	0.004	173.69	-0.026	-0.142	0.09	0.678
SCZ	Neurot	0.038	-0.021	0.095	0.318	0.014	-0.074	0.101	0.767	0.021	62.85	0.024	-0.078	0.126	0.695
SCZ	SCZ	0.025	-0.031	0.081	0.51	0.031	-0.066	0.126	0.536	0.029	-23.5	-0.006	-0.112	0.102	0.925
SCZ	SRH	-0.113	-0.175	-0.05	0.003	-0.084	-0.174	0.003	0.088	-0.092	25.87	-0.029	-0.143	0.085	0.638

Note. BMI = Body Mass Index; IQ = Intelligence; GCSE = General Certificate of Secondary Education (educational achievement); ADHD = Attention-Deficit/Hyperactivity Disorder; SCZ = Schizophrenia symptoms; EA = Educational Attainment; Neurot = Neuroticism; SRH = Self-rated Health; B = Between-family estimate; W = Within-family estimate; L.CI = Lower 95% bootstrap percentile interval; U.CI = Upper 95% bootstrap percentile interval; P = statistical significance of beta estimate; TotEff = Total effect derived as the intra-class correlation weighted sum of the within- and between family effect. PercRed = Reduction of prediction estimates when comparing within- to between-family estimates in percentage. BetaDiff = Difference of between-family and within-family beta coefficient; P.diff = statistical significance of difference between within- and between-family estimates.

Table S15. Within- and between-family prediction estimates for opposite-sex twin pairs

pheno	GPS	beta.B	L.CI.B	U.CI.B	P.B	beta.W	L.CI.W	U.CI.W	P.W	TotEff	PercRed	BetaDiff	Diff.L.CI	Diff.U.CI	P.diff
ADHD	ADHD	0.087	0.046	0.127	0.002	0.081	0.014	0.147	0.017	0.083	7.65	0.007	-0.073	0.087	0.878
ADHD	BMI	0.099	0.058	0.14	2.64e-04	0.029	-0.039	0.097	0.415	0.051	70.61	0.07	-0.009	0.149	0.118
ADHD	EA	-0.114	-0.151	-0.076	1.77e-05	-0.057	-0.133	0.018	0.132	-0.074	50.36	-0.057	-0.143	0.029	0.211
ADHD	Height	0.035	-0.007	0.077	0.197	0.049	-0.014	0.112	0.15	0.044	-38.19	-0.013	-0.089	0.061	0.757
ADHD	IQ	-0.105	-0.146	-0.065	8.45e-05	-0.125	-0.192	-0.057	3.36e-04	-0.119	-18.31	0.019	-0.058	0.098	0.659
ADHD	Neurot	0.046	0.003	0.089	0.109	0.013	-0.05	0.076	0.687	0.023	71.17	0.033	-0.039	0.105	0.453
ADHD	SCZ	-0.017	-0.058	0.023	0.533	0.015	-0.051	0.082	0.659	0.005	189.84	-0.033	-0.109	0.046	0.463
ADHD	SRH	-0.116	-0.156	-0.075	2.15e-05	0.005	-0.064	0.077	0.882	-0.032	104.5	-0.121	-0.203	-0.039	0.006
BMI	ADHD	0.055	-0.004	0.114	0.148	0.094	0.003	0.184	0.049	0.083	-72.02	-0.039	-0.145	0.07	0.517
BMI	BMI	0.375	0.322	0.427	7.22e-26	0.279	0.193	0.366	1.12e-08	0.3	25.66	0.096	-0.004	0.194	0.102
BMI	EA	-0.103	-0.161	-0.044	0.005	-0.057	-0.163	0.048	0.293	-0.07	45.25	-0.047	-0.164	0.072	0.471
BMI	Height	0.011	-0.046	0.066	0.764	0.014	-0.079	0.11	0.757	0.013	-32.91	-0.004	-0.111	0.104	0.952
BMI	IQ	-0.027	-0.079	0.026	0.455	0.008	-0.097	0.11	0.869	-0.002	129.74	-0.035	-0.15	0.081	0.562
BMI	Neurot	-0.113	-0.178	-0.05	0.003	-0.011	-0.096	0.078	0.821	-0.039	90.63	-0.103	-0.213	0.004	0.089
BMI	SCZ	-0.092	-0.146	-0.039	0.015	-0.024	-0.117	0.071	0.618	-0.043	73.52	-0.068	-0.176	0.037	0.272
BMI	SRH	-0.102	-0.164	-0.039	0.008	-0.117	-0.206	-0.029	0.014	-0.113	-14.89	0.015	-0.089	0.122	0.804
GCSE	ADHD	-0.164	-0.207	-0.12	8.52e-08	-0.061	-0.111	-0.01	0.024	-0.12	62.87	-0.103	-0.169	-0.035	0.011
GCSE	BMI	-0.106	-0.151	-0.061	4.84e-04	-0.023	-0.08	0.035	0.424	-0.071	78.62	-0.084	-0.158	-0.009	0.044
GCSE	EA	0.423	0.384	0.461	4.32e-51	0.204	0.147	0.259	4.47e-12	0.32	51.78	0.219	0.152	0.289	3.02e-08
GCSE	Height	0.003	-0.042	0.047	0.91	-0.029	-0.084	0.025	0.277	-0.01	973.95	0.033	-0.038	0.103	0.416
GCSE	IQ	0.305	0.264	0.347	4.22e-25	0.198	0.147	0.25	5.85e-13	0.259	35	0.107	0.04	0.172	0.007
GCSE	Neurot	-0.062	-0.111	-0.015	0.052	-0.009	-0.06	0.042	0.719	-0.04	84.94	-0.053	-0.122	0.015	0.2
GCSE	SCZ	0.098	0.054	0.144	0.001	-0.03	-0.084	0.025	0.278	0.045	130.25	0.128	0.057	0.2	0.002
GCSE	SRH	0.144	0.098	0.191	3.98e-06	0.048	-0.008	0.103	0.084	0.103	66.87	0.096	0.022	0.169	0.02
SRH	ADHD	-0.042	-0.093	0.008	0.206	-0.054	-0.16	0.051	0.298	-0.053	-29.16	0.012	-0.105	0.131	0.843

SRH	BMI	-0.076	-0.129	-0.025	0.019	0.016	-0.084	0.119	0.756	0.009	121.54	-0.093	-0.206	0.02	0.134
SRH	EA	0.075	0.025	0.123	0.02	0.087	-0.022	0.194	0.134	0.086	-16.59	-0.012	-0.132	0.109	0.852
SRH	Height	0.061	0.013	0.113	0.052	-0.029	-0.125	0.071	0.569	-0.021	146.86	0.09	-0.021	0.198	0.129
SRH	IQ	-0.007	-0.06	0.046	0.822	0.008	-0.103	0.114	0.886	0.006	206.96	-0.015	-0.13	0.105	0.812
SRH	Neurot	-0.069	-0.123	-0.014	0.041	-0.001	-0.094	0.09	0.98	-0.007	98.18	-0.067	-0.175	0.039	0.268
SRH	SCZ	-0.003	-0.053	0.046	0.922	-0.048	-0.152	0.055	0.362	-0.044	-1,423.01	0.045	-0.069	0.16	0.466
SRH	SRH	0.144	0.095	0.193	6.68e-06	0.02	-0.078	0.116	0.715	0.02	86.4	0.124	0.015	0.233	0.045
Height	ADHD	-0.047	-0.107	0.015	0.228	-0.046	-0.129	0.038	0.263	-0.046	1.82	-0.001	-0.105	0.101	0.988
Height	BMI	-0.015	-0.074	0.044	0.689	0.064	-0.016	0.14	0.123	0.029	521.91	-0.079	-0.175	0.02	0.158
Height	EA	0.036	-0.021	0.092	0.338	0.032	-0.058	0.123	0.493	0.033	11.83	0.004	-0.095	0.105	0.943
Height	Height	0.5	0.452	0.552	3.83e-49	0.429	0.349	0.509	1.35e-29	0.457	14.26	0.071	-0.028	0.169	0.135
Height	IQ	0.019	-0.042	0.079	0.609	0.039	-0.047	0.125	0.346	0.03	-110.25	-0.021	-0.119	0.08	0.71
Height	Neurot	-0.051	-0.106	0.004	0.19	-0.012	-0.09	0.064	0.764	-0.029	76.38	-0.039	-0.132	0.057	0.485
Height	SCZ	-0.043	-0.1	0.014	0.257	-0.012	-0.089	0.063	0.779	-0.025	72.92	-0.031	-0.125	0.062	0.577
Height	SRH	0.061	0.006	0.118	0.115	-0.03	-0.109	0.046	0.461	0.01	149.07	0.091	-0.004	0.189	0.104
IQ	ADHD	-0.095	-0.153	-0.039	0.008	0.024	-0.055	0.103	0.549	-0.021	125.78	-0.119	-0.216	-0.023	0.028
IQ	BMI	-0.052	-0.105	0	0.143	0.011	-0.071	0.092	0.787	-0.013	121.3	-0.063	-0.158	0.034	0.244
IQ	EA	0.236	0.186	0.287	2.89e-12	0.106	0.016	0.195	0.021	0.153	54.8	0.129	0.027	0.234	0.023
IQ	Height	-0.052	-0.104	0.001	0.144	0.04	-0.045	0.121	0.317	0.004	175.93	-0.092	-0.188	0.006	0.084
IQ	IQ	0.259	0.206	0.312	3.11e-14	0.127	0.048	0.205	0.002	0.174	50.92	0.132	0.034	0.23	0.013
IQ	Neurot	-0.032	-0.089	0.028	0.381	0.048	-0.028	0.122	0.208	0.017	251.49	-0.08	-0.176	0.018	0.128
IQ	SCZ	0.033	-0.021	0.089	0.365	0.021	-0.059	0.101	0.603	0.026	35.04	0.011	-0.084	0.108	0.833
IQ	SRH	0.104	0.051	0.158	0.004	0.019	-0.064	0.103	0.651	0.052	81.28	0.084	-0.015	0.185	0.131
Neurot	ADHD	0.005	-0.065	0.077	0.909	0.106	-0.029	0.238	0.131	0.093	-1,915.36	-0.101	-0.252	0.052	0.228
Neurot	BMI	-0.004	-0.067	0.059	0.923	-0.023	-0.168	0.118	0.742	-0.021	-428.58	0.019	-0.139	0.182	0.822
Neurot	EA	0.011	-0.062	0.086	0.789	-0.082	-0.228	0.068	0.294	-0.071	824.02	0.094	-0.07	0.255	0.292
Neurot	Height	-0.037	-0.108	0.029	0.393	0.055	-0.077	0.188	0.427	0.043	246.22	-0.092	-0.241	0.057	0.257

Neurot	IQ	0.026	-0.043	0.095	0.572	0.09	-0.048	0.233	0.193	0.082	-247.44	-0.064	-0.222	0.086	0.438
Neurot	Neurot	0.087	0.007	0.173	0.076	-0.006	-0.14	0.133	0.932	0.005	106.6	0.093	-0.067	0.256	0.265
Neurot	SCZ	0.013	-0.068	0.092	0.782	0.018	-0.125	0.163	0.791	0.018	-43.33	-0.006	-0.175	0.165	0.947
Neurot	SRH	-0.02	-0.087	0.049	0.667	-0.039	-0.182	0.102	0.579	-0.036	-96.1	0.019	-0.135	0.178	0.82
SCZ	ADHD	0.012	-0.048	0.071	0.773	-0.012	-0.117	0.096	0.823	-0.008	204.09	0.024	-0.101	0.144	0.724
SCZ	BMI	0.099	0.041	0.158	0.01	-0.01	-0.13	0.115	0.861	0.01	110.12	0.109	-0.027	0.242	0.111
SCZ	EA	-0.102	-0.163	-0.039	0.01	0.043	-0.077	0.16	0.499	0.016	142.06	-0.145	-0.281	-0.004	0.052
SCZ	Height	-0.026	-0.092	0.042	0.496	0.026	-0.071	0.123	0.64	0.016	201.05	-0.052	-0.175	0.071	0.44
SCZ	IQ	-0.013	-0.076	0.05	0.731	0.039	-0.068	0.147	0.52	0.029	389.35	-0.052	-0.18	0.074	0.466
SCZ	Neurot	0.082	0.025	0.142	0.043	-0.07	-0.167	0.026	0.181	-0.041	185.91	0.152	0.038	0.267	0.022
SCZ	SCZ	-0.058	-0.117	0.002	0.148	-0.06	-0.176	0.055	0.289	-0.06	-4.98	0.003	-0.131	0.137	0.967
SCZ	SRH	-0.11	-0.167	-0.053	0.006	0.069	-0.044	0.18	0.214	0.035	162.53	-0.179	-0.299	-0.058	0.009

Note. BMI = Body Mass Index; IQ = Intelligence; GCSE = General Certificate of Secondary Education (educational achievement); ADHD = Attention-Deficit/Hyperactivity Disorder; SCZ = Schizophrenia symptoms; EA = Educational Attainment; Neurot = Neuroticism; SRH = Self-rated Health; B = Between-family estimate; W = Within-family estimate; L.CI = Lower 95% bootstrap percentile interval; U.CI = Upper 95% bootstrap percentile interval; P = statistical significance of beta estimate; TotEff = Total effect derived as the intra-class correlation weighted sum of the within- and between family effect. PercRed = Reduction of prediction estimates when comparing within- to between-family estimates in percentage. BetaDiff = Difference of between-family and within-family beta coefficient; P.diff = statistical significance of difference between within- and between-family estimates.

Table S16. Within- and between-family prediction estimates after accounting for family socio-economic status for opposite-sex twin pairs

pheno	GPS	beta.B	L.CI.B	U.CI.B	P.B	beta.W	L.CI.W	U.CI.W	P.W	TotEff	PercRed	BetaDiff	Diff.L.CI	Diff.U.CI	P.diff
ADHD	ADHD	0.056	0.013	0.099	0.042	0.074	0.009	0.137	0.028	0.068	-30.95	-0.017	-0.095	0.063	0.688
ADHD	BMI	0.059	0.018	0.1	0.03	0.039	-0.03	0.11	0.267	0.045	32.97	0.019	-0.062	0.1	0.664
ADHD	EA	-0.034	-0.076	0.008	0.241	-0.055	-0.129	0.022	0.14	-0.049	-61.93	0.021	-0.068	0.109	0.656
ADHD	Height	0.027	-0.016	0.07	0.319	0.05	-0.013	0.114	0.139	0.043	-86.08	-0.023	-0.102	0.052	0.592
ADHD	IQ	-0.064	-0.106	-0.023	0.017	-0.108	-0.17	-0.044	0.002	-0.094	-66.99	0.043	-0.033	0.119	0.323
ADHD	Neurot	0.022	-0.019	0.064	0.429	0.014	-0.048	0.078	0.662	0.017	36.68	0.008	-0.065	0.081	0.848
ADHD	SCZ	-0.013	-0.051	0.027	0.64	0.019	-0.047	0.084	0.593	0.009	246.76	-0.031	-0.106	0.045	0.477
ADHD	SRH	-0.071	-0.111	-0.031	0.011	0.006	-0.062	0.076	0.865	-0.017	108.36	-0.077	-0.158	0.003	0.084
BMI	ADHD	0.038	-0.018	0.095	0.312	0.093	0.002	0.184	0.056	0.079	-144.38	-0.055	-0.163	0.051	0.371
BMI	BMI	0.354	0.301	0.405	2.49e-23	0.277	0.189	0.37	2.77e-08	0.293	21.66	0.077	-0.026	0.179	0.2
BMI	EA	-0.027	-0.089	0.035	0.507	-0.063	-0.171	0.045	0.25	-0.053	-137.05	0.036	-0.092	0.159	0.591
BMI	Height	-0.01	-0.068	0.045	0.773	0.012	-0.08	0.107	0.8	0.006	215.54	-0.022	-0.131	0.085	0.706
BMI	IQ	0.011	-0.041	0.064	0.758	0.013	-0.091	0.12	0.798	0.012	-13.05	-0.001	-0.118	0.115	0.981
BMI	Neurot	-0.11	-0.172	-0.046	0.004	0.014	-0.069	0.1	0.766	-0.017	113.12	-0.124	-0.232	-0.018	0.043
BMI	SCZ	-0.083	-0.135	-0.031	0.026	-0.021	-0.112	0.076	0.683	-0.036	75.26	-0.062	-0.172	0.042	0.318
BMI	SRH	-0.069	-0.135	-0.004	0.08	-0.125	-0.213	-0.038	0.01	-0.11	-81.81	0.056	-0.054	0.166	0.366
GCSE	ADHD	-0.067	-0.106	-0.03	0.011	-0.052	-0.1	-0.002	0.053	-0.059	22.96	-0.015	-0.081	0.048	0.681
GCSE	BMI	-0.023	-0.06	0.014	0.38	-0.026	-0.083	0.032	0.356	-0.025	-15.65	0.004	-0.066	0.071	0.926
GCSE	EA	0.238	0.199	0.276	1.42e-18	0.198	0.142	0.254	2.00e-11	0.216	16.86	0.04	-0.029	0.107	0.309
GCSE	Height	0.002	-0.038	0.041	0.947	-0.033	-0.087	0.022	0.222	-0.017	2,047.97	0.035	-0.034	0.104	0.348
GCSE	IQ	0.199	0.163	0.237	9.77e-15	0.191	0.141	0.244	3.30e-12	0.195	4.07	0.008	-0.057	0.072	0.827
GCSE	Neurot	-0.007	-0.047	0.034	0.8	-0.012	-0.062	0.04	0.656	-0.009	-68.75	0.005	-0.06	0.069	0.9
GCSE	SCZ	0.066	0.028	0.104	0.011	-0.034	-0.088	0.022	0.214	0.012	152.31	0.1	0.031	0.166	0.008
GCSE	SRH	0.025	-0.017	0.068	0.36	0.043	-0.013	0.097	0.122	0.034	-71.17	-0.018	-0.089	0.056	0.647

SRH	ADHD	-0.04	-0.091	0.01	0.216	-0.062	-0.171	0.042	0.255	-0.062	-55.49	0.022	-0.093	0.14	0.726
SRH	BMI	-0.081	-0.134	-0.027	0.011	0.014	-0.094	0.119	0.801	0.014	117.42	-0.095	-0.212	0.025	0.138
SRH	EA	0.072	0.016	0.125	0.036	0.077	-0.031	0.19	0.205	0.077	-7.84	-0.006	-0.131	0.115	0.936
SRH	Height	0.061	0.01	0.113	0.049	-0.03	-0.131	0.072	0.568	-0.03	149.83	0.091	-0.019	0.205	0.136
SRH	IQ	-0.017	-0.071	0.039	0.586	0.007	-0.102	0.114	0.902	0.007	140.75	-0.024	-0.143	0.099	0.709
SRH	Neurot	-0.056	-0.112	0	0.088	-0.004	-0.098	0.088	0.944	-0.004	93.18	-0.052	-0.159	0.055	0.41
SRH	SCZ	-0.01	-0.06	0.038	0.76	-0.044	-0.148	0.061	0.413	-0.042	-345.92	0.034	-0.081	0.149	0.586
SRH	SRH	0.135	0.085	0.185	4.63e-05	0.015	-0.083	0.113	0.778	0.015	88.72	0.12	0.01	0.23	0.058
Height	ADHD	-0.04	-0.104	0.024	0.311	-0.054	-0.138	0.03	0.189	-0.047	-34.38	0.014	-0.089	0.119	0.809
Height	BMI	-0.017	-0.079	0.043	0.658	0.046	-0.034	0.126	0.277	0.017	365.72	-0.063	-0.163	0.038	0.271
Height	EA	0.035	-0.029	0.099	0.401	0.046	-0.049	0.14	0.32	0.041	-30.51	-0.011	-0.12	0.099	0.863
Height	Height	0.519	0.471	0.568	2.98e-51	0.411	0.336	0.488	2.47e-27	0.453	20.67	0.107	0.016	0.2	0.025
Height	IQ	0.012	-0.051	0.074	0.756	0.062	-0.024	0.145	0.141	0.039	-423.67	-0.05	-0.147	0.049	0.376
Height	Neurot	-0.044	-0.101	0.013	0.268	-0.017	-0.096	0.061	0.669	-0.03	60.57	-0.027	-0.122	0.07	0.638
Height	SCZ	-0.047	-0.106	0.012	0.222	-0.015	-0.093	0.062	0.721	-0.029	68.12	-0.032	-0.128	0.064	0.573
Height	SRH	0.055	-0.004	0.114	0.173	-0.016	-0.096	0.058	0.686	0.016	129.83	0.072	-0.024	0.171	0.211
IQ	ADHD	-0.043	-0.1	0.013	0.206	0.029	-0.051	0.11	0.477	0.006	168.25	-0.072	-0.168	0.024	0.175
IQ	BMI	0.017	-0.035	0.066	0.617	-0.004	-0.085	0.078	0.93	0.003	121.98	0.02	-0.076	0.118	0.702
IQ	EA	0.094	0.04	0.148	0.007	0.101	0.01	0.192	0.03	0.098	-7.5	-0.007	-0.114	0.101	0.903
IQ	Height	-0.034	-0.083	0.015	0.305	0.042	-0.042	0.124	0.293	0.018	223.1	-0.076	-0.17	0.02	0.142
IQ	IQ	0.183	0.133	0.233	2.81e-08	0.122	0.041	0.202	0.004	0.14	33.5	0.061	-0.036	0.158	0.245
IQ	Neurot	-0.001	-0.057	0.055	0.981	0.026	-0.05	0.102	0.507	0.017	3,246.29	-0.027	-0.125	0.07	0.605
IQ	SCZ	0.025	-0.027	0.081	0.449	0.016	-0.064	0.099	0.695	0.019	36.99	0.009	-0.09	0.106	0.859
IQ	SRH	0.015	-0.037	0.067	0.656	0.029	-0.055	0.113	0.494	0.025	-91.59	-0.014	-0.112	0.087	0.798
Neurot	ADHD	0.002	-0.071	0.08	0.97	0.106	-0.026	0.24	0.134	0.093	-5,910.41	-0.104	-0.26	0.049	0.218
Neurot	BMI	-0.007	-0.073	0.062	0.882	-0.024	-0.169	0.119	0.736	-0.022	-248.42	0.017	-0.144	0.183	0.84
Neurot	EA	0.027	-0.047	0.101	0.555	-0.061	-0.211	0.093	0.446	-0.051	325.45	0.089	-0.082	0.255	0.338

Neurot	Height	-0.033	-0.105	0.038	0.465	0.055	-0.082	0.189	0.427	0.044	267.33	-0.088	-0.239	0.067	0.285
Neurot	IQ	0.033	-0.037	0.104	0.486	0.118	-0.02	0.258	0.093	0.107	-259.05	-0.085	-0.242	0.072	0.312
Neurot	Neurot	0.085	0.003	0.173	0.091	-0.017	-0.155	0.122	0.807	-0.005	119.7	0.101	-0.063	0.267	0.23
Neurot	SCZ	0.02	-0.061	0.102	0.668	0.014	-0.133	0.159	0.842	0.015	30.54	0.006	-0.165	0.177	0.942
Neurot	SRH	-0.004	-0.075	0.067	0.939	-0.019	-0.168	0.129	0.789	-0.017	-415.93	0.015	-0.148	0.18	0.858
SCZ	ADHD	-0.006	-0.067	0.056	0.891	-0.009	-0.116	0.102	0.873	-0.008	-58.7	0.003	-0.125	0.128	0.962
SCZ	BMI	0.099	0.038	0.159	0.009	-0.015	-0.141	0.115	0.793	0.004	115.37	0.114	-0.027	0.251	0.098
SCZ	EA	-0.055	-0.127	0.018	0.203	0.058	-0.063	0.178	0.366	0.039	205.5	-0.113	-0.257	0.036	0.143
SCZ	Height	-0.055	-0.115	0.008	0.14	0.019	-0.082	0.118	0.734	0.006	134.4	-0.074	-0.197	0.049	0.268
SCZ	IQ	0.022	-0.043	0.089	0.572	0.034	-0.076	0.146	0.573	0.032	-52.99	-0.012	-0.143	0.121	0.87
SCZ	Neurot	0.074	0.018	0.132	0.064	-0.09	-0.18	0.001	0.091	-0.062	221.34	0.164	0.055	0.272	0.013
SCZ	SCZ	-0.04	-0.1	0.02	0.312	-0.073	-0.192	0.044	0.208	-0.068	-84.02	0.033	-0.103	0.17	0.633
SCZ	SRH	-0.088	-0.145	-0.029	0.03	0.082	-0.032	0.194	0.143	0.053	193.45	-0.17	-0.292	-0.048	0.014

Note. BMI = Body Mass Index; IQ = Intelligence; GCSE = General Certificate of Secondary Education (educational achievement); ADHD = Attention-Deficit/Hyperactivity Disorder; SCZ = Schizophrenia symptoms; EA = Educational Attainment; Neurot = Neuroticism; SRH = Self-rated Health; B = Between-family estimate; W = Within-family estimate; L.CI = Lower 95% bootstrap percentile interval; U.CI = Upper 95% bootstrap percentile interval; P = statistical significance of beta estimate; TotEff = Total effect derived as the intra-class correlation weighted sum of the within- and between family effect. PercRed = Reduction of prediction estimates when comparing within- to between-family estimates in percentage. BetaDiff = Difference of between-family and within-family beta coefficient; P.diff = statistical significance of difference between within- and between-family estimates.

Table S17. Within- and between-family prediction estimates based on twin pairs genotyped on OEE chip

pheno	GPS	beta.B	L.CI.B	U.CI.B	P.B	beta.W	L.CI.W	U.CI.W	P.W	TotEff	PercRed	BetaDiff	Diff.L.CI	Diff.U.CI	P.diff
ADHD	ADHD	0.138	0.094	0.183	4.29e-06	0.161	0.088	0.232	8.31e-06	0.154	-16.12	-0.022	-0.109	0.063	6.33e-01
ADHD	BMI	0.114	0.072	0.156	5.34e-05	0.028	-0.042	0.099	4.48e-01	0.054	75.56	0.086	0.004	0.169	6.21e-02
ADHD	EA	-0.159	-0.204	-0.114	3.24e-08	-0.132	-0.213	-0.053	8.50e-04	-0.14	16.66	-0.026	-0.115	0.065	5.87e-01
ADHD	Height	-0.021	-0.065	0.022	4.59e-01	0.028	-0.046	0.102	4.60e-01	0.012	230.8	-0.049	-0.134	0.036	2.99e-01
ADHD	IQ	-0.143	-0.183	-0.102	3.59e-07	-0.131	-0.202	-0.059	4.38e-04	-0.135	7.84	-0.011	-0.094	0.07	8.10e-01
ADHD	Neurot	0.065	0.021	0.11	2.47e-02	0.049	-0.021	0.119	1.79e-01	0.054	25.71	0.017	-0.063	0.095	7.17e-01
ADHD	SCZ	0.015	-0.028	0.059	6.14e-01	-0.01	-0.083	0.06	7.85e-01	-0.002	169.47	0.025	-0.056	0.11	5.99e-01
ADHD	SRH	-0.15	-0.192	-0.106	1.33e-07	0.009	-0.066	0.086	8.06e-01	-0.038	106.21	-0.159	-0.247	-0.071	7.41e-04
BMI	ADHD	0.112	0.057	0.168	4.48e-03	0.137	0.044	0.227	3.51e-03	0.128	-22.37	-0.025	-0.129	0.081	6.81e-01
BMI	BMI	0.35	0.292	0.407	1.54e-20	0.317	0.233	0.402	3.40e-12	0.327	9.31	0.033	-0.069	0.135	5.71e-01
BMI	EA	-0.126	-0.184	-0.066	1.58e-03	-0.088	-0.189	0.011	8.49e-02	-0.1	29.66	-0.037	-0.157	0.081	5.64e-01
BMI	Height	-0.009	-0.065	0.045	8.12e-01	-0.043	-0.15	0.061	3.60e-01	-0.032	-380.55	0.034	-0.085	0.152	5.71e-01
BMI	IQ	-0.049	-0.106	0.01	1.89e-01	-0.023	-0.126	0.076	6.34e-01	-0.032	53.42	-0.026	-0.143	0.096	6.66e-01
BMI	Neurot	-0.088	-0.142	-0.03	2.30e-02	-0.111	-0.196	-0.025	1.97e-02	-0.103	-26.71	0.023	-0.08	0.127	7.01e-01
BMI	SCZ	-0.041	-0.107	0.025	2.97e-01	-0.083	-0.18	0.011	8.84e-02	-0.069	-100.31	0.041	-0.074	0.159	5.07e-01
BMI	SRH	-0.136	-0.196	-0.075	5.31e-04	-0.171	-0.261	-0.082	3.93e-04	-0.159	-25.65	0.035	-0.076	0.145	5.72e-01
GCSE	ADHD	-0.184	-0.231	-0.136	1.29e-08	-0.08	-0.136	-0.024	5.99e-03	-0.138	56.58	-0.104	-0.176	-0.031	1.60e-02
GCSE	BMI	-0.179	-0.225	-0.133	1.21e-08	-0.041	-0.099	0.018	1.58e-01	-0.119	77.01	-0.138	-0.213	-0.063	1.22e-03
GCSE	EA	0.423	0.38	0.465	5.77e-43	0.196	0.138	0.257	1.65e-10	0.314	53.6	0.227	0.151	0.3	8.15e-08
GCSE	Height	0.047	0	0.094	1.24e-01	-0.022	-0.082	0.038	4.62e-01	0.018	146.52	0.069	-0.007	0.146	1.06e-01
GCSE	IQ	0.284	0.241	0.326	2.61e-20	0.211	0.154	0.27	6.88e-13	0.252	25.58	0.073	-0.001	0.146	8.24e-02
GCSE	Neurot	-0.13	-0.181	-0.082	4.12e-05	-0.06	-0.116	-0.005	3.38e-02	-0.1	54.22	-0.071	-0.144	0.003	9.46e-02
GCSE	SCZ	0.058	0.01	0.105	6.87e-02	0.002	-0.061	0.063	9.44e-01	0.034	96.46	0.056	-0.022	0.134	1.97e-01
GCSE	SRH	0.193	0.145	0.24	6.61e-10	0.038	-0.021	0.097	2.00e-01	0.124	80.46	0.155	0.079	0.233	2.73e-04

SRH	ADHD	-0.049	-0.104	0.007	1.70e-01	-0.087	-0.188	0.012	8.50e-02	-0.08	-77.87	0.038	-0.073	0.156	5.37e-01
SRH	BMI	-0.115	-0.166	-0.064	8.35e-04	-0.032	-0.128	0.063	5.11e-01	-0.046	71.74	-0.083	-0.192	0.024	1.69e-01
SRH	EA	0.117	0.062	0.172	8.36e-04	0.049	-0.05	0.147	3.67e-01	0.06	58.57	0.069	-0.042	0.178	2.84e-01
SRH	Height	0.053	0.001	0.106	1.15e-01	0.014	-0.091	0.118	7.76e-01	0.021	72.92	0.039	-0.078	0.155	5.23e-01
SRH	IQ	0.028	-0.026	0.081	4.05e-01	0.042	-0.061	0.146	4.10e-01	0.04	-49.87	-0.014	-0.133	0.102	8.19e-01
SRH	Neurot	-0.084	-0.137	-0.032	1.46e-02	0.016	-0.085	0.113	7.53e-01	-0.001	118.95	-0.1	-0.208	0.015	1.02e-01
SRH	SCZ	-0.07	-0.124	-0.016	4.70e-02	-0.048	-0.146	0.048	3.55e-01	-0.051	31.97	-0.022	-0.131	0.086	7.19e-01
SRH	SRH	0.128	0.071	0.182	2.31e-04	0.115	0.018	0.218	2.38e-02	0.117	9.8	0.013	-0.107	0.126	8.38e-01
Height	ADHD	-0.026	-0.09	0.039	5.03e-01	-0.053	-0.139	0.034	2.22e-01	-0.042	-99.85	0.026	-0.08	0.136	6.51e-01
Height	BMI	-0.018	-0.078	0.043	6.34e-01	-0.027	-0.105	0.052	5.28e-01	-0.023	-47.15	0.009	-0.088	0.106	8.80e-01
Height	EA	0.036	-0.029	0.097	3.60e-01	-0.052	-0.145	0.044	2.69e-01	-0.017	243.91	0.088	-0.028	0.199	1.50e-01
Height	Height	0.472	0.425	0.519	2.38e-42	0.474	0.389	0.564	1.44e-30	0.473	-0.37	-0.002	-0.103	0.096	9.73e-01
Height	IQ	0.062	0.003	0.12	9.24e-02	-0.013	-0.094	0.07	7.70e-01	0.017	120.53	0.075	-0.022	0.172	1.89e-01
Height	Neurot	-0.048	-0.108	0.014	2.03e-01	-0.009	-0.099	0.088	8.42e-01	-0.024	81.95	-0.039	-0.148	0.067	4.93e-01
Height	SCZ	-0.023	-0.08	0.035	5.63e-01	0.052	-0.032	0.138	2.34e-01	0.022	331.46	-0.075	-0.175	0.024	2.02e-01
Height	SRH	0.063	0.008	0.117	1.01e-01	-0.018	-0.098	0.061	6.82e-01	0.014	128.57	0.082	-0.016	0.181	1.64e-01
IQ	ADHD	-0.056	-0.126	0.01	2.01e-01	-0.038	-0.133	0.059	4.29e-01	-0.045	32.09	-0.018	-0.132	0.094	7.82e-01
IQ	BMI	-0.035	-0.095	0.023	3.98e-01	-0.015	-0.107	0.08	7.55e-01	-0.023	57.33	-0.02	-0.131	0.09	7.51e-01
IQ	EA	0.254	0.183	0.328	1.17e-08	0.12	0.018	0.221	2.64e-02	0.169	52.93	0.135	0.011	0.258	5.20e-02
IQ	Height	0.018	-0.051	0.084	6.62e-01	0.111	0.011	0.21	2.14e-02	0.074	-501.79	-0.093	-0.213	0.027	1.47e-01
IQ	IQ	0.237	0.17	0.305	3.03e-08	0.124	0.03	0.216	1.28e-02	0.166	47.44	0.112	-0.003	0.23	8.46e-02
IQ	Neurot	-0.016	-0.083	0.051	7.18e-01	-0.029	-0.129	0.069	5.48e-01	-0.024	-84.6	0.013	-0.105	0.135	8.38e-01
IQ	SCZ	-0.031	-0.094	0.033	4.87e-01	0.006	-0.092	0.103	9.07e-01	-0.009	118.71	-0.036	-0.154	0.08	5.81e-01
IQ	SRH	0.142	0.08	0.207	6.32e-04	0.06	-0.037	0.153	2.47e-01	0.091	57.93	0.082	-0.03	0.196	2.12e-01
Neurot	ADHD	-0.051	-0.124	0.025	3.16e-01	0.051	-0.096	0.193	5.25e-01	0.051	199.75	-0.101	-0.268	0.066	2.81e-01
Neurot	BMI	-0.015	-0.092	0.069	7.52e-01	-0.101	-0.257	0.05	2.06e-01	-0.101	-584.1	0.087	-0.086	0.268	3.49e-01
Neurot	EA	0.102	0.031	0.171	2.64e-02	0.053	-0.139	0.249	5.59e-01	0.053	47.91	0.049	-0.163	0.259	6.30e-01

Neurot	Height	0.001	-0.069	0.072	9.76e-01	-0.069	-0.219	0.079	4.05e-01	-0.069	5,025.39	0.07	-0.099	0.24	4.55e-01
Neurot	IQ	0.081	0.015	0.148	7.68e-02	-0.004	-0.167	0.164	9.59e-01	-0.004	105.04	0.085	-0.09	0.26	3.54e-01
Neurot	Neurot	0.152	0.072	0.237	1.45e-03	-0.037	-0.21	0.14	6.36e-01	-0.037	124.46	0.189	0.005	0.373	3.83e-02
Neurot	SCZ	0.06	-0.023	0.143	2.27e-01	0.118	-0.05	0.286	1.57e-01	0.118	-97.87	-0.059	-0.253	0.137	5.45e-01
Neurot	SRH	-0.025	-0.095	0.048	5.82e-01	-0.02	-0.177	0.135	8.12e-01	-0.02	21.09	-0.005	-0.177	0.169	9.55e-01
SCZ	ADHD	0.067	0.01	0.124	1.22e-01	-0.018	-0.112	0.08	7.50e-01	0.002	126.43	0.084	-0.031	0.198	2.28e-01
SCZ	BMI	0.147	0.081	0.213	3.53e-04	0.044	-0.068	0.162	4.32e-01	0.068	69.79	0.102	-0.028	0.231	1.40e-01
SCZ	EA	-0.098	-0.164	-0.031	2.09e-02	0.053	-0.064	0.174	3.75e-01	0.017	153.99	-0.151	-0.292	-0.011	3.86e-02
SCZ	Height	0.017	-0.051	0.087	6.80e-01	0.03	-0.07	0.138	5.86e-01	0.027	-82.66	-0.014	-0.15	0.12	8.41e-01
SCZ	IQ	-0.006	-0.071	0.059	8.79e-01	0.015	-0.102	0.136	7.94e-01	0.01	352.11	-0.022	-0.158	0.11	7.62e-01
SCZ	Neurot	0.049	-0.011	0.111	2.43e-01	0.032	-0.063	0.13	5.65e-01	0.036	35.46	0.017	-0.097	0.13	8.01e-01
SCZ	SCZ	0.006	-0.057	0.072	8.81e-01	0.072	-0.045	0.188	2.08e-01	0.056	-1,026.73	-0.066	-0.201	0.072	3.56e-01
SCZ	SRH	-0.083	-0.15	-0.019	4.78e-02	-0.036	-0.149	0.075	5.24e-01	-0.047	56.93	-0.047	-0.177	0.085	4.99e-01

Note. BMI = Body Mass Index; IQ = Intelligence; GCSE = General Certificate of Secondary Education (educational achievement); ADHD = Attention-Deficit/Hyperactivity Disorder; SCZ = Schizophrenia symptoms; EA = Educational Attainment; Neurot = Neuroticism; SRH = Self-rated Health; B = Between-family estimate; W = Within-family estimate; L.CI = Lower 95% bootstrap percentile interval; U.CI = Upper 95% bootstrap percentile interval; P = statistical significance of beta estimate; TotEff = Total effect derived as the intra-class correlation weighted sum of the within- and between family effect. PercRed = Reduction of prediction estimates when comparing within- to between-family estimates in percentage. BetaDiff = Difference of between-family and within-family beta coefficient; P.diff = statistical significance of difference between within- and between-family estimates.

Table S18. Within- and between-family prediction estimates after accounting for family socio-economic status based on twin pairs genotyped on OEE chip

pheno	GPS	beta.B	L.CI.B	U.CI.B	P.B	beta.W	L.CI.W	U.CI.W	P.W	TotEff	PercRed	BetaDiff	Diff.L.CI	Diff.U.CI	P.diff
ADHD	ADHD	0.101	0.055	0.149	6.99e-04	0.158	0.088	0.229	9.92e-06	0.141	-56	-0.057	-0.142	0.029	0.221
ADHD	BMI	0.062	0.021	0.104	0.028	0.037	-0.033	0.109	0.314	0.044	40.44	0.025	-0.059	0.107	0.587
ADHD	EA	-0.078	-0.125	-0.03	0.012	-0.121	-0.198	-0.042	0.002	-0.108	-55.69	0.043	-0.048	0.134	0.387
ADHD	Height	-0.012	-0.055	0.031	0.673	0.037	-0.035	0.111	0.31	0.023	419.85	-0.049	-0.134	0.036	0.286
ADHD	IQ	-0.107	-0.148	-0.064	1.34e-04	-0.114	-0.186	-0.042	0.002	-0.112	-6.59	0.007	-0.074	0.088	0.879
ADHD	Neurot	0.046	0.004	0.089	0.103	0.053	-0.018	0.124	0.137	0.051	-15.6	-0.007	-0.087	0.071	0.875
ADHD	SCZ	0.019	-0.024	0.063	0.5	-0.005	-0.078	0.065	0.89	0.002	127.06	0.025	-0.057	0.108	0.604
ADHD	SRH	-0.106	-0.147	-0.065	2.01e-04	0.012	-0.064	0.088	0.753	-0.022	111.09	-0.118	-0.205	-0.032	0.012
BMI	ADHD	0.078	0.024	0.135	0.045	0.128	0.035	0.221	0.007	0.113	-63.62	-0.05	-0.157	0.058	0.416
BMI	BMI	0.329	0.273	0.383	2.40e-18	0.318	0.23	0.403	4.80e-12	0.321	3.29	0.011	-0.088	0.112	0.852
BMI	EA	-0.066	-0.132	-0.001	0.12	-0.085	-0.189	0.012	0.102	-0.079	-28.13	0.019	-0.104	0.146	0.781
BMI	Height	-0.007	-0.062	0.047	0.849	-0.04	-0.145	0.065	0.404	-0.03	-467.72	0.033	-0.085	0.153	0.585
BMI	IQ	-0.028	-0.085	0.031	0.46	-0.011	-0.114	0.09	0.818	-0.016	59.49	-0.016	-0.133	0.103	0.789
BMI	Neurot	-0.101	-0.156	-0.043	0.008	-0.095	-0.183	-0.006	0.055	-0.097	5.84	-0.006	-0.111	0.102	0.924
BMI	SCZ	-0.047	-0.109	0.018	0.231	-0.083	-0.179	0.013	0.097	-0.072	-77.54	0.036	-0.079	0.153	0.566
BMI	SRH	-0.091	-0.153	-0.03	0.022	-0.182	-0.275	-0.094	1.81e-04	-0.154	-100.22	0.091	-0.022	0.203	0.144
GCSE	ADHD	-0.08	-0.122	-0.037	0.005	-0.077	-0.133	-0.024	0.008	-0.078	3.45	-0.003	-0.071	0.068	0.946
GCSE	BMI	-0.086	-0.125	-0.048	0.002	-0.042	-0.102	0.018	0.15	-0.062	51.11	-0.044	-0.118	0.028	0.273
GCSE	EA	0.254	0.209	0.297	8.21e-18	0.183	0.122	0.243	2.56e-09	0.214	28.06	0.071	-0.004	0.146	0.089
GCSE	Height	0.028	-0.014	0.071	0.287	-0.031	-0.089	0.029	0.305	-0.004	208.08	0.059	-0.013	0.132	0.14
GCSE	IQ	0.184	0.144	0.222	1.16e-11	0.205	0.145	0.263	3.06e-12	0.195	-11.16	-0.021	-0.092	0.052	0.603
GCSE	Neurot	-0.083	-0.125	-0.04	0.002	-0.066	-0.122	-0.009	0.021	-0.074	21.15	-0.018	-0.089	0.054	0.656
GCSE	SCZ	0.045	0.005	0.084	0.101	0.005	-0.059	0.067	0.875	0.023	89.62	0.04	-0.034	0.115	0.318
GCSE	SRH	0.064	0.02	0.11	0.023	0.03	-0.027	0.088	0.299	0.046	52.41	0.034	-0.041	0.109	0.408

SRH	ADHD	-0.03	-0.087	0.025	0.399	-0.1	-0.199	-0.001	0.05	-0.089	-229.18	0.07	-0.048	0.186	0.264
SRH	BMI	-0.109	-0.162	-0.056	0.002	-0.03	-0.129	0.069	0.548	-0.043	72.3	-0.079	-0.191	0.033	0.199
SRH	EA	0.103	0.041	0.164	0.008	0.048	-0.049	0.146	0.378	0.057	53.22	0.055	-0.061	0.17	0.412
SRH	Height	0.052	-0.001	0.105	0.126	0.021	-0.084	0.127	0.676	0.026	58.73	0.03	-0.088	0.147	0.619
SRH	IQ	0.012	-0.043	0.069	0.72	0.049	-0.054	0.153	0.339	0.043	-296.33	-0.037	-0.157	0.083	0.552
SRH	Neurot	-0.079	-0.132	-0.027	0.024	0.021	-0.079	0.123	0.695	0.004	126.1	-0.099	-0.213	0.011	0.114
SRH	SCZ	-0.064	-0.121	-0.009	0.074	-0.038	-0.137	0.065	0.48	-0.042	41.3	-0.026	-0.142	0.084	0.68
SRH	SRH	0.121	0.064	0.18	8.23e-04	0.112	0.013	0.211	0.031	0.113	7.8	0.009	-0.11	0.125	0.88
Height	ADHD	-0.015	-0.081	0.053	0.713	-0.057	-0.149	0.03	0.195	-0.04	-284.59	0.042	-0.066	0.153	0.478
Height	BMI	-0.017	-0.08	0.046	0.675	-0.034	-0.116	0.046	0.43	-0.027	-105.03	0.017	-0.082	0.119	0.765
Height	EA	0.022	-0.051	0.09	0.609	-0.055	-0.147	0.042	0.247	-0.024	347.92	0.077	-0.044	0.193	0.229
Height	Height	0.475	0.427	0.522	3.53e-42	0.474	0.391	0.564	1.20e-29	0.475	0.16	0.001	-0.102	0.099	0.988
Height	IQ	0.056	-0.004	0.114	0.143	-0.017	-0.1	0.064	0.7	0.012	130.5	0.074	-0.026	0.171	0.21
Height	Neurot	-0.052	-0.114	0.013	0.177	0.002	-0.097	0.103	0.967	-0.019	103.62	-0.054	-0.167	0.06	0.364
Height	SCZ	-0.018	-0.078	0.041	0.656	0.055	-0.033	0.144	0.232	0.026	405.27	-0.073	-0.179	0.031	0.232
Height	SRH	0.059	0.002	0.115	0.143	-0.02	-0.1	0.063	0.663	0.012	132.98	0.079	-0.021	0.176	0.191
IQ	ADHD	-0.034	-0.103	0.034	0.428	-0.028	-0.124	0.072	0.565	-0.03	16.05	-0.005	-0.124	0.109	0.933
IQ	BMI	0.017	-0.04	0.074	0.678	-0.023	-0.117	0.073	0.641	-0.009	236.05	0.039	-0.071	0.149	0.531
IQ	EA	0.149	0.076	0.226	0.001	0.11	0.004	0.215	0.044	0.123	26.37	0.039	-0.089	0.168	0.578
IQ	Height	0.028	-0.034	0.092	0.486	0.104	0.005	0.204	0.032	0.078	-272.58	-0.076	-0.196	0.042	0.225
IQ	IQ	0.18	0.114	0.247	1.51e-05	0.11	0.016	0.206	0.03	0.133	38.8	0.07	-0.05	0.192	0.283
IQ	Neurot	-0.009	-0.074	0.059	0.82	-0.073	-0.175	0.027	0.139	-0.051	-679.52	0.064	-0.058	0.189	0.32
IQ	SCZ	-0.019	-0.082	0.043	0.66	0.01	-0.089	0.108	0.834	0.001	156.29	-0.029	-0.146	0.088	0.657
IQ	SRH	0.081	0.017	0.143	0.046	0.071	-0.022	0.165	0.171	0.074	12.97	0.011	-0.102	0.123	0.872
Neurot	ADHD	-0.053	-0.128	0.03	0.305	0.047	-0.098	0.19	0.563	0.047	187.39	-0.1	-0.268	0.071	0.295
Neurot	BMI	-0.019	-0.101	0.068	0.694	-0.097	-0.255	0.065	0.232	-0.097	-405.99	0.078	-0.105	0.259	0.408
Neurot	EA	0.122	0.047	0.196	0.016	0.053	-0.139	0.243	0.565	0.053	56.7	0.069	-0.145	0.281	0.507

Neurot	Height	0.003	-0.07	0.075	0.953	-0.056	-0.206	0.097	0.505	-0.056	2,095.34	0.058	-0.118	0.231	0.54
Neurot	IQ	0.092	0.02	0.165	0.056	0.015	-0.149	0.183	0.85	0.015	83.43	0.077	-0.101	0.257	0.411
Neurot	Neurot	0.157	0.076	0.242	0.001	-0.058	-0.231	0.114	0.463	-0.058	137.08	0.215	0.031	0.401	0.02
Neurot	SCZ	0.056	-0.027	0.139	0.266	0.1	-0.071	0.272	0.237	0.1	-79.51	-0.044	-0.244	0.151	0.65
Neurot	SRH	-0.03	-0.105	0.044	0.522	-0.011	-0.172	0.142	0.9	-0.011	64.87	-0.02	-0.191	0.157	0.839
SCZ	ADHD	0.046	-0.013	0.106	0.287	-0.016	-0.114	0.084	0.778	-0.002	134.23	0.062	-0.056	0.18	0.381
SCZ	BMI	0.147	0.078	0.214	4.31e-04	0.051	-0.063	0.17	0.365	0.072	64.97	0.095	-0.038	0.224	0.174
SCZ	EA	-0.076	-0.153	0.002	0.104	0.059	-0.059	0.179	0.332	0.028	176.8	-0.135	-0.279	0.01	0.077
SCZ	Height	0.021	-0.046	0.091	0.61	0.035	-0.071	0.141	0.532	0.032	-71.84	-0.015	-0.151	0.122	0.831
SCZ	IQ	0.017	-0.051	0.086	0.69	0.012	-0.106	0.132	0.841	0.013	27.32	0.005	-0.131	0.14	0.95
SCZ	Neurot	0.033	-0.027	0.097	0.429	0.006	-0.083	0.1	0.911	0.012	80.94	0.027	-0.082	0.135	0.701
SCZ	SCZ	0.009	-0.056	0.076	0.837	0.055	-0.067	0.17	0.352	0.045	-523.79	-0.047	-0.181	0.097	0.525
SCZ	SRH	-0.06	-0.13	0.008	0.165	-0.026	-0.133	0.082	0.648	-0.034	56.92	-0.034	-0.164	0.095	0.63

Note. BMI = Body Mass Index; IQ = Intelligence; GCSE = General Certificate of Secondary Education (educational achievement); ADHD = Attention-Deficit/Hyperactivity Disorder; SCZ = Schizophrenia symptoms; EA = Educational Attainment; Neurot = Neuroticism; SRH = Self-rated Health; B = Between-family estimate; W = Within-family estimate; L.CI = Lower 95% bootstrap percentile interval; U.CI = Upper 95% bootstrap percentile interval; P = statistical significance of beta estimate; TotEff = Total effect derived as the intra-class correlation weighted sum of the within- and between family effect. PercRed = Reduction of prediction estimates when comparing within- to between-family estimates in percentage. BetaDiff = Difference of between-family and within-family beta coefficient; P.diff = statistical significance of difference between within- and between-family estimates.

Table S19. Within- and between-family prediction estimates based on twin pairs with one twin genotyped on the OEE chip and the co-twin genotyped on the Affymetrix chip

pheno	GPS	beta.B	L.CI.B	U.CI.B	P.B	beta.W	L.CI.W	U.CI.W	P.W	TotEff	PercRed	BetaDiff	Diff.L.CI	Diff.U.CI	P.diff
ADHD	ADHD	0.093	0.055	0.131	1.73e-04	0.093	0.032	0.154	2.72e-03	0.093	-0.33	0	-0.07	0.07	9.94e-01
ADHD	BMI	0.074	0.037	0.112	3.08e-03	0.046	-0.014	0.106	1.45e-01	0.055	38.76	0.029	-0.041	0.099	4.71e-01
ADHD	EA	-0.088	-0.124	-0.052	2.44e-04	-0.147	-0.216	-0.078	8.16e-06	-0.127	-66.99	0.059	-0.018	0.136	1.46e-01
ADHD	Height	-0.016	-0.054	0.021	5.08e-01	0.01	-0.048	0.067	7.52e-01	0.001	159.15	-0.026	-0.094	0.041	5.08e-01
ADHD	IQ	-0.073	-0.111	-0.035	3.24e-03	-0.114	-0.174	-0.053	4.35e-04	-0.1	-55.06	0.04	-0.031	0.112	3.21e-01
ADHD	Neurot	0.028	-0.011	0.066	2.70e-01	0.056	-0.003	0.114	6.23e-02	0.047	-97.59	-0.027	-0.096	0.041	4.83e-01
ADHD	SCZ	-0.008	-0.046	0.029	7.33e-01	0.076	0.013	0.136	1.47e-02	0.048	999.37	-0.084	-0.154	-0.013	3.35e-02
ADHD	SRH	-0.104	-0.139	-0.068	3.22e-05	-0.072	-0.133	-0.011	2.61e-02	-0.083	30.68	-0.032	-0.1	0.037	4.35e-01
BMI	ADHD	0.032	-0.02	0.083	3.46e-01	-0.012	-0.099	0.071	7.81e-01	0.001	138.03	0.044	-0.054	0.142	4.25e-01
BMI	BMI	0.351	0.304	0.398	9.42e-28	0.279	0.196	0.362	8.94e-11	0.297	20.35	0.071	-0.022	0.165	1.73e-01
BMI	EA	-0.103	-0.157	-0.053	1.81e-03	-0.048	-0.143	0.046	3.04e-01	-0.064	53.52	-0.055	-0.161	0.049	3.33e-01
BMI	Height	-0.079	-0.13	-0.029	1.58e-02	-0.022	-0.103	0.059	6.16e-01	-0.038	72.36	-0.057	-0.151	0.038	2.93e-01
BMI	IQ	-0.042	-0.093	0.008	2.11e-01	0.04	-0.046	0.129	3.69e-01	0.016	194.82	-0.083	-0.186	0.017	1.41e-01
BMI	Neurot	-0.072	-0.124	-0.019	4.05e-02	0.078	0	0.156	6.28e-02	0.033	208.71	-0.149	-0.245	-0.055	5.99e-03
BMI	SCZ	-0.055	-0.105	-0.006	1.00e-01	-0.033	-0.116	0.05	4.45e-01	-0.04	40.35	-0.022	-0.118	0.073	6.83e-01
BMI	SRH	-0.15	-0.206	-0.097	1.14e-05	-0.102	-0.185	-0.022	1.97e-02	-0.116	32.1	-0.048	-0.144	0.047	3.83e-01
GCSE	ADHD	-0.172	-0.212	-0.132	7.47e-10	-0.045	-0.095	0.005	7.56e-02	-0.119	73.61	-0.127	-0.191	-0.062	7.78e-04
GCSE	BMI	-0.096	-0.138	-0.054	8.21e-04	-0.007	-0.056	0.042	7.77e-01	-0.059	92.48	-0.089	-0.154	-0.025	2.05e-02
GCSE	EA	0.416	0.381	0.45	1.42e-57	0.228	0.176	0.281	9.28e-18	0.33	45.12	0.188	0.123	0.251	1.90e-07
GCSE	Height	0.02	-0.022	0.061	4.84e-01	0.016	-0.036	0.068	5.23e-01	0.018	18.23	0.004	-0.063	0.07	9.24e-01
GCSE	IQ	0.35	0.312	0.388	5.54e-37	0.192	0.143	0.243	1.72e-13	0.281	44.98	0.157	0.095	0.22	2.26e-05
GCSE	Neurot	-0.027	-0.072	0.017	3.49e-01	-0.062	-0.111	-0.013	1.03e-02	-0.041	-126.36	0.034	-0.032	0.102	3.61e-01
GCSE	SCZ	0.022	-0.021	0.063	4.39e-01	-0.032	-0.082	0.018	2.05e-01	0	248.09	0.053	-0.012	0.118	1.54e-01
GCSE	SRH	0.183	0.143	0.224	1.61e-10	0.074	0.02	0.128	4.30e-03	0.137	59.58	0.109	0.042	0.178	4.50e-03

SRH	ADHD	-0.062	-0.108	-0.016	3.94e-02	-0.059	-0.149	0.029	2.10e-01	-0.059	5.07	-0.003	-0.101	0.097	9.55e-01
SRH	BMI	-0.1	-0.146	-0.056	7.62e-04	0.008	-0.078	0.094	8.69e-01	-0.003	107.79	-0.108	-0.203	-0.013	5.27e-02
SRH	EA	0.034	-0.009	0.079	2.39e-01	0.104	0.009	0.201	3.84e-02	0.096	-204.18	-0.07	-0.173	0.037	2.28e-01
SRH	Height	0.069	0.024	0.115	1.60e-02	-0.027	-0.114	0.06	5.65e-01	-0.017	138.84	0.096	0.001	0.195	7.93e-02
SRH	IQ	0.017	-0.031	0.064	5.59e-01	0.014	-0.086	0.113	7.75e-01	0.014	19.5	0.003	-0.106	0.113	9.53e-01
SRH	Neurot	-0.077	-0.124	-0.029	1.17e-02	-0.037	-0.118	0.045	4.15e-01	-0.041	52.31	-0.04	-0.135	0.054	4.58e-01
SRH	SCZ	-0.023	-0.07	0.023	4.30e-01	-0.031	-0.122	0.062	5.04e-01	-0.03	-35.01	0.008	-0.098	0.113	8.83e-01
SRH	SRH	0.146	0.102	0.189	9.70e-07	0.043	-0.053	0.139	3.66e-01	0.053	70.44	0.103	-0.002	0.207	6.68e-02
Height	ADHD	-0.042	-0.096	0.012	2.26e-01	-0.06	-0.135	0.013	1.02e-01	-0.052	-42.09	0.018	-0.074	0.112	7.25e-01
Height	BMI	-0.022	-0.075	0.031	5.32e-01	0.068	-0.003	0.139	6.26e-02	0.026	416.18	-0.09	-0.18	0.003	7.40e-02
Height	EA	0.032	-0.019	0.083	3.42e-01	0.033	-0.044	0.112	3.99e-01	0.033	-2.99	-0.001	-0.092	0.088	9.85e-01
Height	Height	0.46	0.415	0.507	2.76e-49	0.361	0.292	0.427	1.19e-24	0.402	21.58	0.099	0.015	0.186	2.65e-02
Height	IQ	0.024	-0.029	0.077	4.92e-01	-0.001	-0.078	0.074	9.79e-01	0.01	104.17	0.025	-0.066	0.119	6.30e-01
Height	Neurot	-0.068	-0.118	-0.017	5.76e-02	0.018	-0.051	0.087	6.01e-01	-0.022	127.16	-0.086	-0.17	0	8.52e-02
Height	SCZ	-0.072	-0.126	-0.02	3.29e-02	-0.012	-0.082	0.057	7.31e-01	-0.04	82.79	-0.06	-0.144	0.026	2.26e-01
Height	SRH	0.052	-0.002	0.106	1.34e-01	-0.034	-0.105	0.035	3.63e-01	0.006	164.38	0.086	-0.001	0.174	9.03e-02
IQ	ADHD	-0.099	-0.144	-0.056	7.52e-04	0.013	-0.05	0.079	6.89e-01	-0.035	113.44	-0.112	-0.19	-0.034	1.13e-02
IQ	BMI	-0.025	-0.069	0.02	4.04e-01	0.029	-0.036	0.094	3.73e-01	0.006	216.45	-0.055	-0.133	0.025	2.22e-01
IQ	EA	0.255	0.213	0.296	1.17e-19	0.13	0.059	0.202	2.74e-04	0.181	49.06	0.125	0.043	0.207	5.40e-03
IQ	Height	0.004	-0.042	0.051	8.99e-01	0.011	-0.056	0.078	7.26e-01	0.008	-198.99	-0.008	-0.089	0.073	8.64e-01
IQ	IQ	0.275	0.233	0.317	1.26e-20	0.143	0.076	0.208	2.99e-05	0.197	48.04	0.132	0.056	0.213	3.10e-03
IQ	Neurot	-0.02	-0.068	0.03	5.06e-01	-0.013	-0.077	0.051	6.86e-01	-0.016	36.28	-0.007	-0.089	0.076	8.68e-01
IQ	SCZ	0.022	-0.025	0.067	4.69e-01	0.015	-0.049	0.08	6.37e-01	0.018	28.88	0.006	-0.071	0.085	8.88e-01
IQ	SRH	0.091	0.045	0.137	2.41e-03	0.004	-0.067	0.073	9.06e-01	0.042	95.55	0.087	0.003	0.171	5.56e-02
Neurot	ADHD	0.014	-0.046	0.078	7.16e-01	0.102	-0.002	0.209	8.70e-02	0.087	-608.24	-0.088	-0.21	0.033	2.19e-01
Neurot	BMI	-0.016	-0.077	0.047	7.07e-01	0.101	-0.017	0.216	7.93e-02	0.081	746.62	-0.116	-0.249	0.019	9.93e-02
Neurot	EA	-0.033	-0.09	0.026	3.78e-01	-0.01	-0.133	0.119	8.75e-01	-0.014	68.2	-0.022	-0.161	0.109	7.68e-01

Neurot	Height	-0.041	-0.106	0.023	3.02e-01	0.1	-0.009	0.209	8.10e-02	0.076	345.2	-0.141	-0.266	-0.017	4.24e-02
Neurot	IQ	-0.06	-0.121	0.002	1.41e-01	0.107	-0.013	0.225	9.26e-02	0.079	277.73	-0.167	-0.299	-0.034	2.65e-02
Neurot	Neurot	0.076	0.01	0.14	7.40e-02	0.061	-0.045	0.161	2.90e-01	0.063	19.46	0.015	-0.105	0.138	8.36e-01
Neurot	SCZ	-0.055	-0.12	0.012	1.66e-01	0.028	-0.092	0.145	6.34e-01	0.015	151.84	-0.083	-0.218	0.056	2.44e-01
Neurot	SRH	-0.047	-0.105	0.011	2.35e-01	-0.027	-0.153	0.097	6.46e-01	-0.031	41.67	-0.02	-0.157	0.119	7.84e-01
SCZ	ADHD	0.027	-0.029	0.083	4.62e-01	0.019	-0.083	0.124	6.98e-01	0.021	30.3	0.008	-0.114	0.131	8.93e-01
SCZ	BMI	0.08	0.031	0.129	2.49e-02	0.037	-0.061	0.134	4.48e-01	0.048	54.14	0.043	-0.061	0.147	4.70e-01
SCZ	EA	-0.071	-0.127	-0.012	4.44e-02	-0.025	-0.125	0.072	6.29e-01	-0.037	64.27	-0.045	-0.156	0.071	4.70e-01
SCZ	Height	-0.013	-0.068	0.039	7.03e-01	-0.02	-0.106	0.064	6.79e-01	-0.018	-54.38	0.007	-0.092	0.104	9.05e-01
SCZ	IQ	-0.038	-0.094	0.017	2.77e-01	0.032	-0.062	0.128	5.18e-01	0.014	184.3	-0.071	-0.185	0.041	2.47e-01
SCZ	Neurot	0.071	0.015	0.126	5.63e-02	-0.062	-0.147	0.025	1.71e-01	-0.027	186.78	0.133	0.032	0.237	2.27e-02
SCZ	SCZ	-0.021	-0.073	0.032	5.57e-01	-0.071	-0.166	0.024	1.40e-01	-0.058	-239.65	0.05	-0.057	0.159	4.02e-01
SCZ	SRH	-0.152	-0.205	-0.096	3.39e-05	0.006	-0.087	0.1	8.98e-01	-0.034	104.07	-0.158	-0.266	-0.051	9.00e-03

Note. BMI = Body Mass Index; IQ = Intelligence; GCSE = General Certificate of Secondary Education (educational achievement); ADHD = Attention-Deficit/Hyperactivity Disorder; SCZ = Schizophrenia symptoms; EA = Educational Attainment; Neurot = Neuroticism; SRH = Self-rated Health; B = Between-family estimate; W = Within-family estimate; L.CI = Lower 95% bootstrap percentile interval; U.CI = Upper 95% bootstrap percentile interval; P = statistical significance of beta estimate; TotEff = Total effect derived as the intra-class correlation weighted sum of the within- and between family effect. PercRed = Reduction of prediction estimates when comparing within- to between-family estimates in percentage. BetaDiff = Difference of between-family and within-family beta coefficient; P.diff = statistical significance of difference between within- and between-family estimates.

Table S20. Within- and between-family prediction estimates after accounting for family socio-economic status based on twin pairs with one twin genotyped on the OEE chip and the co-twin genotyped on the Affymetrix chip

pheno	GPS	beta.B	L.CI.B	U.CI.B	P.B	beta.W	L.CI.W	U.CI.W	P.W	TotEff	PercRed	BetaDiff	Diff.L.CI	Diff.U.CI	P.diff
ADHD	ADHD	0.08	0.041	0.119	0.001	0.093	0.029	0.154	0.003	0.089	-15.98	-0.013	-0.084	0.06	0.751
ADHD	BMI	0.05	0.012	0.088	0.044	0.044	-0.017	0.106	0.165	0.046	12.18	0.006	-0.063	0.077	0.879
ADHD	EA	-0.032	-0.072	0.008	0.215	-0.153	-0.224	-0.082	4.30e-06	-0.116	-377.76	0.121	0.041	0.203	0.004
ADHD	Height	-0.012	-0.05	0.024	0.614	0.009	-0.049	0.067	0.779	0.002	170.33	-0.021	-0.09	0.046	0.594
ADHD	IQ	-0.033	-0.073	0.007	0.198	-0.111	-0.171	-0.051	6.93e-04	-0.087	-237.51	0.078	0.006	0.151	0.059
ADHD	Neurot	0.019	-0.019	0.057	0.441	0.056	-0.003	0.116	0.063	0.045	-190	-0.037	-0.106	0.032	0.348
ADHD	SCZ	-0.006	-0.044	0.03	0.794	0.07	0.008	0.132	0.025	0.047	1,210.46	-0.077	-0.147	-0.005	0.053
ADHD	SRH	-0.068	-0.105	-0.032	0.007	-0.07	-0.131	-0.007	0.034	-0.069	-1.73	0.001	-0.07	0.071	0.977
BMI	ADHD	0.012	-0.038	0.06	0.727	-0.013	-0.1	0.072	0.771	-0.006	209.58	0.025	-0.073	0.123	0.657
BMI	BMI	0.33	0.282	0.378	2.87e-25	0.281	0.199	0.366	1.08e-10	0.293	14.69	0.048	-0.047	0.143	0.357
BMI	EA	-0.022	-0.079	0.032	0.532	-0.044	-0.14	0.05	0.347	-0.038	-98.2	0.022	-0.089	0.133	0.71
BMI	Height	-0.08	-0.13	-0.031	0.013	-0.021	-0.103	0.06	0.622	-0.038	73.22	-0.059	-0.154	0.035	0.278
BMI	IQ	0.008	-0.043	0.058	0.824	0.037	-0.049	0.126	0.409	0.029	-387.09	-0.03	-0.132	0.07	0.601
BMI	Neurot	-0.07	-0.122	-0.018	0.042	0.08	0.002	0.158	0.058	0.039	214.12	-0.15	-0.246	-0.056	0.006
BMI	SCZ	-0.048	-0.098	0.001	0.149	-0.03	-0.114	0.052	0.488	-0.035	36.85	-0.018	-0.112	0.078	0.747
BMI	SRH	-0.114	-0.169	-0.062	0.001	-0.108	-0.192	-0.026	0.014	-0.11	5.94	-0.007	-0.103	0.09	0.903
GCSE	ADHD	-0.087	-0.122	-0.051	2.71e-04	-0.049	-0.099	0.002	0.06	-0.066	43.45	-0.038	-0.101	0.026	0.285
GCSE	BMI	-0.031	-0.067	0.007	0.207	-0.008	-0.058	0.042	0.752	-0.018	72.92	-0.022	-0.084	0.039	0.531
GCSE	EA	0.236	0.201	0.27	6.18e-22	0.233	0.181	0.286	6.88e-18	0.234	1.02	0.002	-0.062	0.068	0.946
GCSE	Height	-0.01	-0.046	0.026	0.686	0.02	-0.033	0.073	0.445	0.006	303.31	-0.029	-0.092	0.035	0.403
GCSE	IQ	0.236	0.203	0.269	2.28e-22	0.195	0.144	0.246	2.71e-13	0.212	17.36	0.041	-0.019	0.1	0.249
GCSE	Neurot	0.002	-0.034	0.037	0.95	-0.061	-0.111	-0.011	0.013	-0.033	4,064.67	0.062	-0.001	0.124	0.072
GCSE	SCZ	0.015	-0.021	0.051	0.514	-0.035	-0.086	0.016	0.167	-0.013	332.33	0.051	-0.01	0.112	0.144
GCSE	SRH	0.074	0.037	0.111	0.003	0.075	0.021	0.13	0.004	0.075	-1.97	-0.001	-0.067	0.064	0.968

SRH	ADHD	-0.058	-0.105	-0.012	0.054	-0.053	-0.144	0.035	0.261	-0.054	8.56	-0.005	-0.103	0.096	0.929
SRH	BMI	-0.102	-0.147	-0.058	5.82e-04	0	-0.088	0.089	0.997	-0.009	100.15	-0.102	-0.2	-0.005	0.068
SRH	EA	0.016	-0.032	0.064	0.609	0.102	0.004	0.2	0.045	0.093	-531.24	-0.086	-0.192	0.023	0.151
SRH	Height	0.063	0.017	0.11	0.029	-0.031	-0.119	0.056	0.513	-0.022	148.77	0.094	-0.003	0.193	0.088
SRH	IQ	0.003	-0.048	0.052	0.931	0.014	-0.087	0.114	0.779	0.013	-425.23	-0.011	-0.123	0.1	0.847
SRH	Neurot	-0.068	-0.116	-0.019	0.027	-0.043	-0.126	0.04	0.346	-0.045	36.79	-0.025	-0.122	0.071	0.649
SRH	SCZ	-0.024	-0.071	0.022	0.405	-0.041	-0.134	0.052	0.382	-0.039	-68.94	0.017	-0.09	0.121	0.762
SRH	SRH	0.134	0.089	0.179	1.25e-05	0.044	-0.053	0.142	0.358	0.052	66.96	0.09	-0.017	0.196	0.115
Height	ADHD	-0.035	-0.09	0.021	0.326	-0.064	-0.139	0.009	0.081	-0.05	-84.18	0.029	-0.063	0.124	0.565
Height	BMI	-0.015	-0.07	0.038	0.656	0.053	-0.018	0.125	0.148	0.02	442.54	-0.069	-0.16	0.023	0.174
Height	EA	0.007	-0.05	0.063	0.858	0.041	-0.036	0.119	0.298	0.024	-516.46	-0.034	-0.13	0.06	0.525
Height	Height	0.474	0.431	0.518	5.33e-51	0.341	0.275	0.406	3.37e-22	0.395	28.05	0.133	0.054	0.213	0.003
Height	IQ	0.002	-0.053	0.058	0.956	0.021	-0.055	0.095	0.586	0.012	-941.6	-0.019	-0.109	0.077	0.72
Height	Neurot	-0.059	-0.111	-0.008	0.098	0.015	-0.055	0.083	0.674	-0.02	124.8	-0.074	-0.16	0.013	0.139
Height	SCZ	-0.078	-0.132	-0.026	0.022	-0.019	-0.09	0.051	0.605	-0.047	76.14	-0.06	-0.144	0.026	0.23
Height	SRH	0.038	-0.018	0.094	0.293	-0.02	-0.091	0.05	0.595	0.008	151.51	0.058	-0.031	0.146	0.264
IQ	ADHD	-0.044	-0.086	-0.003	0.107	0.01	-0.053	0.077	0.756	-0.009	123.91	-0.054	-0.134	0.026	0.209
IQ	BMI	0.024	-0.017	0.065	0.392	0.025	-0.041	0.091	0.455	0.025	-4.7	-0.001	-0.077	0.078	0.979
IQ	EA	0.12	0.076	0.164	2.41e-05	0.123	0.051	0.196	6.75e-04	0.122	-2.09	-0.003	-0.087	0.081	0.956
IQ	Height	-0.024	-0.066	0.02	0.39	0.01	-0.058	0.078	0.752	-0.002	143.03	-0.034	-0.113	0.047	0.424
IQ	IQ	0.196	0.157	0.237	4.74e-12	0.139	0.071	0.204	6.57e-05	0.159	29.31	0.058	-0.02	0.138	0.196
IQ	Neurot	0.007	-0.038	0.053	0.805	-0.01	-0.075	0.056	0.76	-0.004	242.7	0.017	-0.065	0.098	0.694
IQ	SCZ	0.019	-0.024	0.061	0.503	0.012	-0.053	0.077	0.712	0.015	34.25	0.006	-0.072	0.084	0.883
IQ	SRH	0.006	-0.038	0.05	0.84	0.002	-0.069	0.073	0.952	0.003	63.87	0.004	-0.079	0.087	0.935
Neurot	ADHD	0.012	-0.054	0.076	0.772	0.109	0.002	0.217	0.072	0.093	-834.85	-0.097	-0.223	0.029	0.179
Neurot	BMI	-0.016	-0.08	0.048	0.703	0.094	-0.024	0.208	0.108	0.076	687.41	-0.11	-0.246	0.029	0.125
Neurot	EA	-0.019	-0.08	0.044	0.631	0.005	-0.121	0.137	0.94	0.001	126.44	-0.024	-0.167	0.114	0.756

Neurot	Height	-0.036	-0.103	0.029	0.374	0.092	-0.019	0.203	0.117	0.071	355.59	-0.128	-0.258	-0.001	0.072
Neurot	IQ	-0.059	-0.124	0.005	0.158	0.122	0.002	0.241	0.057	0.094	307.4	-0.181	-0.317	-0.048	0.018
Neurot	Neurot	0.073	0.007	0.141	0.088	0.063	-0.046	0.172	0.286	0.065	13.89	0.01	-0.12	0.139	0.889
Neurot	SCZ	-0.051	-0.118	0.015	0.198	0.022	-0.099	0.139	0.712	0.011	143.55	-0.073	-0.209	0.068	0.309
Neurot	SRH	-0.032	-0.092	0.028	0.437	-0.017	-0.145	0.109	0.787	-0.019	48.9	-0.016	-0.156	0.128	0.83
SCZ	ADHD	0.017	-0.038	0.073	0.634	0.022	-0.081	0.127	0.66	0.021	-25.13	-0.004	-0.128	0.119	0.943
SCZ	BMI	0.068	0.019	0.116	0.052	0.036	-0.064	0.135	0.466	0.044	47.2	0.032	-0.074	0.139	0.595
SCZ	EA	-0.032	-0.092	0.031	0.385	-0.022	-0.123	0.076	0.675	-0.025	31.05	-0.01	-0.124	0.109	0.877
SCZ	Height	-0.026	-0.075	0.023	0.449	-0.032	-0.119	0.054	0.521	-0.03	-23.63	0.006	-0.092	0.101	0.919
SCZ	IQ	-0.01	-0.066	0.047	0.783	0.028	-0.068	0.126	0.578	0.019	386.8	-0.038	-0.153	0.075	0.539
SCZ	Neurot	0.072	0.017	0.125	0.049	-0.065	-0.153	0.024	0.152	-0.033	191.26	0.137	0.035	0.242	0.019
SCZ	SCZ	-0.012	-0.064	0.04	0.728	-0.068	-0.164	0.028	0.164	-0.055	-458.41	0.056	-0.053	0.165	0.352
SCZ	SRH	-0.131	-0.185	-0.076	3.30e-04	0.011	-0.083	0.106	0.82	-0.021	108.5	-0.142	-0.251	-0.034	0.019

Note. BMI = Body Mass Index; IQ = Intelligence; GCSE = General Certificate of Secondary Education (educational achievement); ADHD = Attention-Deficit/Hyperactivity Disorder; SCZ = Schizophrenia symptoms; EA = Educational Attainment; Neurot = Neuroticism; SRH = Self-rated Health; B = Between-family estimate; W = Within-family estimate; L.CI = Lower 95% bootstrap percentile interval; U.CI = Upper 95% bootstrap percentile interval; P = statistical significance of beta estimate; TotEff = Total effect derived as the intra-class correlation weighted sum of the within- and between family effect. PercRed = Reduction of prediction estimates when comparing within- to between-family estimates in percentage. BetaDiff = Difference of between-family and within-family beta coefficient; P.diff = statistical significance of difference between within- and between-family estimates.

Table S21. Within- and between-family coefficients estimated using a SNP-kinship matrix as random effect

pheno	GPS	beta.B	L.CI.B	U.CI.B	P.B	beta.W	L.CI.W	U.CI.W	P.W	PercRed	BetaDiff	P.diff
ADHD	ADHD	0.113	0.076	0.15	2.52e-09	0.126	0.078	0.174	2.47e-07	-11.74	-0.013	0.668
ADHD	BMI	0.096	0.059	0.132	2.42e-07	0.036	-0.012	0.084	0.143	62.35	0.06	0.053
ADHD	GCSE	-0.114	-0.15	-0.079	2.97e-10	-0.141	-0.192	-0.089	7.91e-08	-23.2	0.027	0.406
ADHD	Height	-0.018	-0.054	0.018	0.334	0.019	-0.029	0.067	0.438	207.18	-0.037	0.23
ADHD	IQ	-0.103	-0.139	-0.067	2.36e-08	-0.127	-0.177	-0.078	4.93e-07	-23.78	0.024	0.435
ADHD	Neurot	0.042	0.004	0.08	0.029	0.048	0.001	0.096	0.046	-14.42	-0.006	0.844
ADHD	SCZ	-0.004	-0.041	0.033	0.835	0.043	-0.006	0.092	0.085	1,189.47	-0.047	0.134
ADHD	SRH	-0.128	-0.164	-0.092	5.21e-12	-0.034	-0.084	0.016	0.182	73.41	-0.094	0.003
BMI	ADHD	0.069	0.018	0.119	0.008	0.054	-0.01	0.117	0.101	21.96	0.015	0.717
BMI	BMI	0.356	0.31	0.402	5.22e-52	0.297	0.236	0.359	2.45e-21	16.47	0.059	0.134
BMI	GCSE	-0.116	-0.166	-0.067	4.59e-06	-0.06	-0.129	0.009	0.088	48.52	-0.056	0.193
BMI	Height	-0.05	-0.099	-0.002	0.04	-0.028	-0.092	0.036	0.389	44.33	-0.022	0.584
BMI	IQ	-0.051	-0.101	-0.002	0.04	0.01	-0.056	0.075	0.771	118.81	-0.061	0.142
BMI	Neurot	-0.085	-0.136	-0.034	0.001	-0.013	-0.076	0.05	0.687	84.8	-0.072	0.081
BMI	SCZ	-0.054	-0.104	-0.004	0.036	-0.054	-0.119	0.01	0.098	-0.64	0	0.993
BMI	SRH	-0.145	-0.195	-0.094	1.61e-08	-0.139	-0.203	-0.074	2.55e-05	4.11	-0.006	0.887
GCSE	ADHD	-0.173	-0.213	-0.134	1.41e-17	-0.059	-0.1	-0.019	0.004	65.73	-0.114	8.69e-05
GCSE	BMI	-0.138	-0.177	-0.098	9.19e-12	-0.014	-0.054	0.027	0.515	90.16	-0.124	1.84e-05
GCSE	GCSE	0.409	0.373	0.445	1.38e-109	0.211	0.169	0.253	5.70e-23	48.5	0.198	1.91e-12
GCSE	Height	0.034	-0.004	0.073	0.082	-0.001	-0.042	0.04	0.978	101.67	0.035	0.225
GCSE	IQ	0.318	0.28	0.355	1.06e-60	0.194	0.153	0.235	2.27e-20	38.97	0.124	1.42e-05
GCSE	Neurot	-0.073	-0.114	-0.032	4.69e-04	-0.057	-0.097	-0.018	0.005	21.31	-0.016	0.594
GCSE	SCZ	0.046	0.006	0.086	0.023	-0.022	-0.063	0.019	0.293	147.92	0.068	0.02
GCSE	SRH	0.189	0.15	0.229	7.00e-21	0.053	0.012	0.094	0.012	72	0.136	2.93e-06
Height	ADHD	-0.033	-0.084	0.018	0.203	-0.059	-0.115	-0.002	0.041	-77.84	0.026	0.507

Height	BMI	-0.021	-0.071	0.029	0.406	0.022	-0.034	0.077	0.45	202.35	-0.043	0.264
Height	GCSE	0.037	-0.013	0.087	0.148	-0.006	-0.067	0.054	0.837	117.31	0.043	0.282
Height	Height	0.464	0.422	0.506	2.75e-104	0.414	0.362	0.466	5.83e-55	10.83	0.05	0.14
Height	IQ	0.042	-0.007	0.091	0.09	-0.001	-0.058	0.057	0.984	101.43	0.043	0.265
Height	Neurot	-0.062	-0.113	-0.011	0.018	0.009	-0.046	0.065	0.743	115.07	-0.071	0.065
Height	SCZ	-0.049	-0.099	0.002	0.057	0.021	-0.036	0.077	0.478	141.99	-0.069	0.073
Height	SRH	0.059	0.009	0.109	0.022	-0.025	-0.082	0.033	0.402	141.69	0.083	0.032
IQ	ADHD	-0.087	-0.135	-0.04	2.82e-04	-0.012	-0.068	0.044	0.67	86.07	-0.075	0.044
IQ	BMI	-0.027	-0.074	0.021	0.269	0.019	-0.036	0.075	0.498	172.19	-0.046	0.218
IQ	GCSE	0.256	0.211	0.301	6.76e-29	0.124	0.064	0.185	5.82e-05	51.5	0.132	6.11e-04
IQ	Height	0.009	-0.038	0.056	0.716	0.044	-0.011	0.099	0.12	-397.34	-0.035	0.346
IQ	IQ	0.267	0.221	0.313	6.56e-30	0.137	0.079	0.194	3.25e-06	48.76	0.13	5.41e-04
IQ	Neurot	-0.016	-0.064	0.032	0.51	-0.016	-0.071	0.039	0.564	0.65	0	0.998
IQ	SCZ	0.009	-0.039	0.057	0.722	0.008	-0.048	0.065	0.768	3.38	0	0.994
IQ	SRH	0.115	0.068	0.162	1.82e-06	0.023	-0.035	0.082	0.437	79.74	0.091	0.017
Neurot	ADHD	-0.012	-0.073	0.049	0.699	0.079	-0.015	0.173	0.1	758.76	-0.091	0.111
Neurot	BMI	-0.024	-0.085	0.037	0.439	0.02	-0.073	0.112	0.677	181.28	-0.044	0.438
Neurot	GCSE	0.015	-0.042	0.071	0.608	0.015	-0.09	0.121	0.778	-2.46	0	0.995
Neurot	Height	-0.022	-0.081	0.037	0.465	0.033	-0.06	0.127	0.483	252.96	-0.055	0.326
Neurot	IQ	-0.014	-0.073	0.046	0.648	0.057	-0.041	0.154	0.254	508.95	-0.071	0.226
Neurot	Neurot	0.11	0.048	0.172	4.68e-04	0.024	-0.068	0.116	0.612	78.4	0.086	0.126
Neurot	SCZ	-0.011	-0.072	0.049	0.714	0.066	-0.029	0.162	0.174	689.82	-0.078	0.178
Neurot	SRH	-0.041	-0.1	0.018	0.172	-0.024	-0.12	0.072	0.622	40.71	-0.017	0.772
SCZ	ADHD	0.041	-0.013	0.096	0.139	-0.004	-0.077	0.069	0.912	109.97	0.045	0.331
SCZ	BMI	0.111	0.059	0.164	3.04e-05	0.035	-0.038	0.109	0.348	68.34	0.076	0.099
SCZ	GCSE	-0.075	-0.128	-0.023	0.005	0.014	-0.064	0.093	0.719	119.22	-0.09	0.063
SCZ	Height	-0.003	-0.054	0.048	0.903	0.005	-0.069	0.078	0.897	253.92	-0.008	0.861

SCZ	IQ	-0.023	-0.074	0.029	0.394	0.031	-0.045	0.107	0.425	237.89	-0.054	0.255
SCZ	Neurot	0.057	0.003	0.112	0.04	-0.024	-0.094	0.046	0.505	141.73	0.081	0.074
SCZ	SCZ	-0.017	-0.071	0.037	0.537	-0.011	-0.085	0.063	0.773	35.51	-0.006	0.898
SCZ	SRH	-0.12	-0.174	-0.067	1.00e-05	-0.008	-0.082	0.065	0.827	93.18	-0.112	0.016
SRH	ADHD	-0.059	-0.103	-0.014	0.011	-0.072	-0.139	-0.004	0.038	-22.6	0.013	0.75
SRH	BMI	-0.107	-0.151	-0.063	1.72e-06	-0.01	-0.078	0.057	0.765	90.41	-0.097	0.018
SRH	GCSE	0.072	0.028	0.116	0.001	0.082	0.009	0.154	0.027	-13.03	-0.009	0.828
SRH	Height	0.063	0.02	0.106	0.004	-0.008	-0.076	0.059	0.813	113.03	0.071	0.082
SRH	IQ	0.018	-0.026	0.061	0.425	0.028	-0.041	0.098	0.421	-60.54	-0.011	0.797
SRH	Neurot	-0.081	-0.126	-0.036	4.19e-04	-0.017	-0.083	0.049	0.618	79.14	-0.064	0.118
SRH	SCZ	-0.044	-0.088	0	0.052	-0.036	-0.104	0.032	0.303	18.49	-0.008	0.845
SRH	SRH	0.138	0.094	0.182	7.53e-10	0.075	0.006	0.144	0.033	45.74	0.063	0.129

Note. BMI = Body Mass Index; IQ = Intelligence; GCSE = General Certificate of Secondary Education (educational achievement); ADHD = Attention-Deficit/Hyperactivity Disorder; SCZ = Schizophrenia symptoms; EA = Educational Attainment; Neurot = Neuroticism; SRH = Self-rated Health; B = Between-family estimate; W = Within-family estimate; L.CI = Lower 95% confidence interval; U.CI = Upper 95% confidence interval; P = statistical significance of beta estimate; PercRed = Reduction of prediction estimates when comparing within- to between-family estimates in percentage. BetaDiff = Difference of between-family and within-family beta coefficient; P.diff = statistical significance of difference between within- and between-family estimates.

Table S22. Within- and between-family prediction estimates using polygenic scores calculated with causal fraction of markers of 0.1

pheno	GPS	beta.B	L.CI.B	U.CI.B	P.B	beta.W	L.CI.W	U.CI.W	P.W	TotEff	PercRed	BetaDiff	Diff.L.CI	Diff.U.CI	P.diff
ADHD	ADHD	0.111	0.082	0.14	6.51e-09	0.127	0.081	0.173	8.05e-08	0.122	-14.24	-0.016	-0.071	0.038	0.602
ADHD	BMI	0.064	0.035	0.094	7.18e-04	-0.007	-0.053	0.039	0.766	0.016	110.93	0.072	0.016	0.125	0.019
ADHD	EA	-0.077	-0.107	-0.049	3.41e-05	-0.104	-0.151	-0.057	1.71e-05	-0.095	-34.31	0.027	-0.028	0.081	0.384
ADHD	Height	-0.003	-0.033	0.027	0.864	0.028	-0.016	0.071	0.228	0.018	960.64	-0.031	-0.085	0.023	0.297
ADHD	IQ	-0.078	-0.107	-0.048	3.25e-05	-0.046	-0.09	-0.002	0.05	-0.056	41.41	-0.032	-0.084	0.022	0.28
ADHD	Neurot	0.008	-0.02	0.037	0.658	0.054	0.009	0.099	0.021	0.039	-538.54	-0.045	-0.098	0.007	0.131
ADHD	SCZ	0.011	-0.018	0.04	0.574	0.005	-0.042	0.052	0.846	0.007	56.68	0.006	-0.05	0.062	0.843
ADHD	SRH	-0.111	-0.139	-0.083	3.63e-09	-0.024	-0.072	0.022	0.322	-0.052	78.27	-0.087	-0.141	-0.03	0.005
BMI	ADHD	0.068	0.031	0.105	0.008	0.07	0.008	0.135	0.027	0.07	-3.17	-0.002	-0.075	0.07	0.958
BMI	BMI	0.23	0.189	0.269	3.05e-19	0.191	0.132	0.251	9.05e-10	0.203	17.13	0.039	-0.03	0.11	0.323
BMI	EA	-0.066	-0.107	-0.025	0.011	-0.066	-0.133	0.002	0.051	-0.066	0.16	0	-0.081	0.081	0.998
BMI	Height	-0.036	-0.074	0.002	0.152	-0.005	-0.07	0.057	0.875	-0.015	86.42	-0.031	-0.105	0.042	0.436
BMI	IQ	-0.01	-0.051	0.031	0.69	0.017	-0.045	0.079	0.591	0.009	270.31	-0.028	-0.1	0.045	0.503
BMI	Neurot	-0.037	-0.079	0.003	0.152	0.015	-0.046	0.077	0.655	-0.002	139.26	-0.052	-0.125	0.021	0.214
BMI	SCZ	-0.051	-0.093	-0.009	0.047	-0.059	-0.121	0.003	0.07	-0.057	-15.07	0.008	-0.065	0.081	0.852
BMI	SRH	-0.115	-0.153	-0.076	8.67e-06	-0.065	-0.127	-0.004	0.041	-0.081	43.23	-0.05	-0.123	0.022	0.226
GCSE	ADHD	-0.176	-0.207	-0.145	6.95e-17	-0.065	-0.102	-0.028	7.34e-04	-0.129	63.08	-0.111	-0.16	-0.063	9.27e-05
GCSE	BMI	-0.088	-0.12	-0.056	4.40e-05	-0.001	-0.039	0.036	0.958	-0.052	98.86	-0.087	-0.137	-0.037	0.002
GCSE	EA	0.29	0.261	0.321	1.05e-45	0.144	0.105	0.184	1.77e-13	0.226	50.39	0.146	0.097	0.197	1.58e-07
GCSE	Height	0.022	-0.01	0.054	0.308	-0.012	-0.05	0.025	0.506	0.007	157.96	0.034	-0.015	0.083	0.229
GCSE	IQ	0.189	0.158	0.22	3.49e-19	0.116	0.077	0.154	8.69e-10	0.158	38.59	0.073	0.022	0.124	0.01
GCSE	Neurot	-0.033	-0.065	0	0.12	-0.04	-0.078	-0.003	0.032	-0.036	-20.67	0.007	-0.043	0.057	0.809
GCSE	SCZ	0.018	-0.014	0.05	0.4	-0.004	-0.04	0.032	0.833	0.009	122.32	0.022	-0.026	0.07	0.442
GCSE	SRH	0.161	0.129	0.193	3.96e-14	0.038	0	0.078	0.05	0.108	76.35	0.123	0.071	0.173	1.87e-05

SRH	ADHD	-0.057	-0.093	-0.023	0.012	-0.071	-0.137	-0.005	0.039	-0.069	-23.81	0.014	-0.062	0.087	0.74
SRH	BMI	-0.043	-0.077	-0.01	0.06	-0.024	-0.087	0.041	0.484	-0.026	45.31	-0.02	-0.091	0.052	0.631
SRH	EA	0.037	0.001	0.072	0.106	0.024	-0.045	0.092	0.511	0.026	35.18	0.013	-0.063	0.09	0.762
SRH	Height	0.052	0.018	0.086	0.022	0.008	-0.056	0.074	0.816	0.014	84.87	0.044	-0.031	0.117	0.279
SRH	IQ	-0.006	-0.04	0.028	0.788	0.011	-0.056	0.079	0.757	0.008	277.46	-0.017	-0.093	0.057	0.685
SRH	Neurot	-0.057	-0.091	-0.024	0.012	0.009	-0.059	0.074	0.804	0	115.2	-0.066	-0.14	0.006	0.115
SRH	SCZ	-0.015	-0.05	0.021	0.518	-0.038	-0.106	0.031	0.282	-0.035	-156.45	0.023	-0.055	0.101	0.581
SRH	SRH	0.128	0.093	0.163	1.35e-08	0.054	-0.01	0.121	0.119	0.063	58.13	0.075	-0.002	0.147	0.069
Height	ADHD	-0.032	-0.074	0.008	0.217	-0.06	-0.115	-0.004	0.032	-0.048	-87.94	0.028	-0.041	0.098	0.461
Height	BMI	-0.039	-0.078	0.001	0.141	0.023	-0.027	0.073	0.397	-0.004	160.27	-0.062	-0.124	0.001	0.103
Height	EA	-0.011	-0.051	0.029	0.672	-0.04	-0.097	0.018	0.179	-0.027	-263	0.029	-0.041	0.099	0.463
Height	Height	0.364	0.329	0.4	4.19e-49	0.304	0.255	0.353	4.21e-30	0.329	16.53	0.06	-0.001	0.122	0.088
Height	IQ	0.052	0.014	0.09	0.044	0.026	-0.029	0.083	0.354	0.038	49.6	0.026	-0.041	0.093	0.5
Height	Neurot	0.004	-0.037	0.045	0.868	0.047	-0.006	0.099	0.105	0.028	-968.57	-0.042	-0.107	0.022	0.277
Height	SCZ	-0.07	-0.11	-0.031	0.007	0.006	-0.049	0.061	0.827	-0.027	108.86	-0.077	-0.146	-0.007	0.047
Height	SRH	0.073	0.035	0.111	0.005	-0.023	-0.075	0.027	0.402	0.019	132.05	0.097	0.031	0.162	0.011
IQ	ADHD	-0.089	-0.126	-0.053	2.45e-04	-0.005	-0.057	0.048	0.868	-0.04	94.87	-0.085	-0.149	-0.019	0.021
IQ	BMI	-0.027	-0.065	0.01	0.281	0.006	-0.046	0.06	0.807	-0.008	123.89	-0.034	-0.1	0.032	0.358
IQ	EA	0.169	0.133	0.206	4.58e-12	0.048	-0.008	0.106	0.089	0.098	71.53	0.121	0.052	0.189	0.001
IQ	Height	0.006	-0.03	0.044	0.797	0.011	-0.042	0.062	0.686	0.009	-63.42	-0.004	-0.067	0.06	0.91
IQ	IQ	0.165	0.13	0.202	1.35e-11	0.101	0.047	0.154	1.55e-04	0.128	38.9	0.064	-0.001	0.131	0.074
IQ	Neurot	0.016	-0.023	0.054	0.536	0.006	-0.048	0.059	0.823	0.01	62.09	0.01	-0.056	0.076	0.791
IQ	SCZ	-0.025	-0.06	0.01	0.312	-0.01	-0.062	0.041	0.718	-0.016	60.36	-0.015	-0.076	0.048	0.683
IQ	SRH	0.101	0.065	0.138	3.89e-05	-0.001	-0.057	0.055	0.973	0.042	100.93	0.102	0.037	0.168	0.006
Neurot	ADHD	-0.014	-0.061	0.034	0.66	0.092	0.005	0.179	0.057	0.081	774.09	-0.105	-0.205	-0.005	0.066
Neurot	BMI	0.013	-0.036	0.063	0.684	-0.064	-0.16	0.029	0.177	-0.056	609.68	0.077	-0.031	0.185	0.175
Neurot	EA	0.012	-0.033	0.059	0.671	-0.067	-0.163	0.031	0.188	-0.058	645.34	0.079	-0.029	0.187	0.175

Neurot	Height	-0.072	-0.119	-0.026	0.017	0.069	-0.02	0.157	0.135	0.055	195.12	-0.141	-0.242	-0.041	0.01
Neurot	IQ	-0.02	-0.067	0.025	0.502	0.1	0.005	0.195	0.037	0.087	589.25	-0.12	-0.224	-0.016	0.034
Neurot	Neurot	0.022	-0.03	0.073	0.482	0.118	0.029	0.207	0.013	0.108	-435.19	-0.096	-0.199	0.008	0.092
Neurot	SCZ	0.013	-0.038	0.063	0.656	-0.04	-0.139	0.057	0.428	-0.034	394.84	0.053	-0.057	0.165	0.363
Neurot	SRH	-0.063	-0.108	-0.018	0.038	-0.003	-0.095	0.091	0.952	-0.009	95.43	-0.06	-0.165	0.044	0.289
SCZ	ADHD	0.05	0.01	0.092	0.069	0.008	-0.063	0.08	0.821	0.019	83.64	0.042	-0.041	0.126	0.357
SCZ	BMI	0.072	0.032	0.113	0.01	0.017	-0.059	0.095	0.632	0.031	75.9	0.055	-0.033	0.139	0.231
SCZ	EA	-0.06	-0.101	-0.018	0.03	0.009	-0.067	0.084	0.828	-0.009	114.36	-0.068	-0.155	0.017	0.155
SCZ	Height	-0.002	-0.048	0.045	0.943	0.008	-0.056	0.072	0.825	0.005	511.19	-0.01	-0.087	0.068	0.826
SCZ	IQ	-0.012	-0.053	0.029	0.662	0.04	-0.04	0.121	0.284	0.027	436.81	-0.052	-0.142	0.038	0.26
SCZ	Neurot	0.047	0.007	0.088	0.088	-0.007	-0.082	0.066	0.851	0.007	114.84	0.054	-0.025	0.136	0.243
SCZ	SCZ	-0.018	-0.06	0.023	0.523	-0.017	-0.089	0.054	0.641	-0.017	1.98	0	-0.081	0.08	0.994
SCZ	SRH	-0.096	-0.137	-0.056	5.37e-04	-0.007	-0.078	0.062	0.838	-0.03	92.35	-0.089	-0.17	-0.007	0.051

Note. BMI = Body Mass Index; IQ = Intelligence; GCSE = General Certificate of Secondary Education (educational achievement); ADHD = Attention-Deficit/Hyperactivity Disorder; SCZ = Schizophrenia symptoms; EA = Educational Attainment; Neurot = Neuroticism; SRH = Self-rated Health; B = Between-family estimate; W = Within-family estimate; L.CI = Lower 95% bootstrap percentile interval; U.CI = Upper 95% bootstrap percentile interval; P = statistical significance of beta estimate; TotEff = Total effect derived as the intra-class correlation weighted sum of the within- and between family effect. PercRed = Reduction of prediction estimates when comparing within- to between-family estimates in percentage. BetaDiff = Difference of between-family and within-family beta coefficient; P.diff = statistical significance of difference between within- and between-family estimates.

Table S23. Within- and between-family prediction estimates after accounting for family socio-economic status using polygenic scores calculated with causal fraction of markers of 0.1

pheno	GPS	beta.B	L.CI.B	U.CI.B	P.B	beta.W	L.CI.W	U.CI.W	P.W	TotEff	PercRed	BetaDiff	Diff.L.CI	Diff.U.CI	P.diff
ADHD	ADHD	0.087	0.057	0.116	5.51e-06	0.125	0.08	0.172	1.26e-07	0.114	-44.9	-0.039	-0.095	0.015	0.2
ADHD	BMI	0.043	0.014	0.073	0.022	-0.005	-0.051	0.041	0.848	0.01	110.7	0.048	-0.006	0.104	0.117
ADHD	EA	-0.031	-0.06	0	0.111	-0.102	-0.15	-0.054	2.83e-05	-0.08	-233.97	0.071	0.017	0.127	0.021
ADHD	Height	0.002	-0.027	0.032	0.904	0.029	-0.014	0.073	0.206	0.021	-1,192.9	-0.027	-0.081	0.027	0.364
ADHD	IQ	-0.059	-0.087	-0.03	0.002	-0.036	-0.081	0.008	0.124	-0.043	38.43	-0.022	-0.075	0.031	0.452
ADHD	Neurot	0.003	-0.026	0.031	0.874	0.054	0.008	0.1	0.021	0.039	-1,722.07	-0.051	-0.105	0.002	0.087
ADHD	SCZ	0.014	-0.014	0.043	0.438	0.001	-0.047	0.049	0.973	0.005	94.43	0.014	-0.043	0.069	0.652
ADHD	SRH	-0.079	-0.108	-0.05	2.93e-05	-0.021	-0.069	0.027	0.387	-0.039	73.34	-0.058	-0.115	-0.001	0.06
BMI	ADHD	0.041	0.006	0.078	0.102	0.063	0.002	0.126	0.049	0.057	-53.54	-0.022	-0.094	0.049	0.588
BMI	BMI	0.211	0.17	0.25	1.08e-16	0.192	0.132	0.253	1.37e-09	0.197	9.01	0.019	-0.053	0.09	0.636
BMI	EA	-0.012	-0.054	0.028	0.64	-0.056	-0.122	0.013	0.104	-0.043	-350.26	0.043	-0.039	0.124	0.316
BMI	Height	-0.036	-0.072	0.001	0.146	-0.004	-0.068	0.061	0.904	-0.013	89.43	-0.032	-0.105	0.043	0.422
BMI	IQ	0.005	-0.036	0.046	0.848	0.016	-0.046	0.078	0.623	0.013	-230.22	-0.011	-0.082	0.063	0.786
BMI	Neurot	-0.038	-0.079	0.002	0.133	0.014	-0.05	0.077	0.68	-0.001	135.89	-0.052	-0.128	0.023	0.214
BMI	SCZ	-0.041	-0.083	0.001	0.112	-0.057	-0.121	0.006	0.087	-0.052	-40.29	0.016	-0.058	0.092	0.696
BMI	SRH	-0.078	-0.118	-0.038	0.003	-0.072	-0.135	-0.012	0.026	-0.074	8.75	-0.007	-0.079	0.066	0.868
GCSE	ADHD	-0.081	-0.109	-0.054	7.87e-06	-0.066	-0.103	-0.03	6.22e-04	-0.073	18.31	-0.015	-0.061	0.031	0.575
GCSE	BMI	-0.044	-0.072	-0.017	0.017	-0.003	-0.042	0.035	0.857	-0.022	92.17	-0.041	-0.089	0.007	0.126
GCSE	EA	0.157	0.13	0.185	9.64e-18	0.142	0.103	0.18	6.83e-13	0.149	9.51	0.015	-0.033	0.064	0.577
GCSE	Height	0.003	-0.025	0.03	0.872	-0.011	-0.049	0.027	0.564	-0.005	476.03	0.014	-0.034	0.062	0.596
GCSE	IQ	0.131	0.105	0.158	3.95e-13	0.116	0.078	0.155	1.04e-09	0.123	11.63	0.015	-0.032	0.063	0.559
GCSE	Neurot	-0.014	-0.042	0.013	0.436	-0.039	-0.078	-0.001	0.04	-0.028	-175.09	0.025	-0.023	0.072	0.345
GCSE	SCZ	0.003	-0.024	0.031	0.869	-0.008	-0.045	0.029	0.674	-0.003	370.11	0.011	-0.034	0.056	0.674
GCSE	SRH	0.063	0.035	0.091	7.41e-04	0.041	0.001	0.079	0.038	0.051	35.18	0.022	-0.026	0.071	0.413

SRH	ADHD	-0.047	-0.082	-0.012	0.042	-0.074	-0.142	-0.008	0.033	-0.071	-58.05	0.027	-0.049	0.103	0.513
SRH	BMI	-0.04	-0.075	-0.006	0.081	-0.027	-0.09	0.037	0.426	-0.029	32.6	-0.013	-0.088	0.059	0.749
SRH	EA	0.021	-0.016	0.058	0.368	0.022	-0.048	0.093	0.543	0.022	-5.01	-0.001	-0.08	0.077	0.981
SRH	Height	0.045	0.012	0.08	0.046	0.008	-0.058	0.075	0.811	0.013	82	0.037	-0.038	0.112	0.364
SRH	IQ	-0.022	-0.057	0.013	0.348	0.013	-0.054	0.081	0.707	0.009	161.5	-0.035	-0.112	0.04	0.408
SRH	Neurot	-0.058	-0.092	-0.024	0.012	0.01	-0.059	0.078	0.786	0.001	116.98	-0.067	-0.144	0.008	0.114
SRH	SCZ	-0.01	-0.045	0.026	0.668	-0.044	-0.113	0.027	0.222	-0.039	-344.75	0.034	-0.045	0.113	0.425
SRH	SRH	0.123	0.087	0.159	1.46e-07	0.055	-0.014	0.122	0.114	0.063	55.18	0.068	-0.009	0.144	0.105
Height	ADHD	-0.022	-0.064	0.02	0.409	-0.064	-0.12	-0.005	0.024	-0.045	-193.3	0.042	-0.03	0.112	0.275
Height	BMI	-0.039	-0.08	0.003	0.144	0.018	-0.031	0.068	0.507	-0.007	147.44	-0.057	-0.12	0.006	0.136
Height	EA	-0.031	-0.073	0.012	0.262	-0.036	-0.094	0.021	0.223	-0.034	-19.3	0.006	-0.067	0.078	0.884
Height	Height	0.371	0.335	0.407	6.64e-50	0.304	0.254	0.353	1.05e-29	0.332	18.12	0.067	0.006	0.13	0.059
Height	IQ	0.044	0.005	0.084	0.094	0.035	-0.022	0.089	0.222	0.039	21.01	0.009	-0.057	0.078	0.811
Height	Neurot	0.008	-0.033	0.048	0.767	0.052	-0.003	0.107	0.075	0.032	-560.59	-0.044	-0.111	0.023	0.263
Height	SCZ	-0.075	-0.116	-0.034	0.004	-0.002	-0.058	0.053	0.938	-0.034	97	-0.073	-0.142	-0.004	0.062
Height	SRH	0.068	0.027	0.109	0.011	-0.016	-0.067	0.035	0.573	0.021	123.34	0.084	0.019	0.149	0.031
IQ	ADHD	-0.042	-0.078	-0.007	0.067	-0.003	-0.058	0.051	0.901	-0.017	91.8	-0.039	-0.104	0.029	0.284
IQ	BMI	0.002	-0.033	0.038	0.916	0.008	-0.046	0.062	0.762	0.006	-231.07	-0.006	-0.072	0.061	0.873
IQ	EA	0.081	0.046	0.116	6.30e-04	0.048	-0.01	0.105	0.099	0.059	41.48	0.034	-0.034	0.102	0.366
IQ	Height	-0.003	-0.037	0.032	0.909	0.003	-0.051	0.056	0.912	0.001	208.87	-0.006	-0.07	0.058	0.874
IQ	IQ	0.125	0.091	0.158	8.18e-08	0.101	0.047	0.154	2.04e-04	0.109	19.38	0.024	-0.04	0.088	0.497
IQ	Neurot	0.019	-0.016	0.055	0.431	-0.004	-0.06	0.053	0.887	0.004	120.82	0.022	-0.045	0.088	0.532
IQ	SCZ	-0.027	-0.059	0.006	0.248	-0.01	-0.064	0.043	0.73	-0.016	64.3	-0.017	-0.079	0.048	0.633
IQ	SRH	0.038	0.002	0.073	0.109	-0.004	-0.059	0.051	0.881	0.01	111.18	0.042	-0.024	0.107	0.253
Neurot	ADHD	-0.016	-0.066	0.033	0.603	0.095	0.006	0.183	0.051	0.083	677.25	-0.112	-0.214	-0.009	0.054
Neurot	BMI	0.013	-0.037	0.062	0.69	-0.056	-0.153	0.042	0.245	-0.049	550.24	0.069	-0.043	0.18	0.232
Neurot	EA	0.02	-0.027	0.069	0.504	-0.062	-0.16	0.036	0.225	-0.054	408.27	0.083	-0.026	0.19	0.165

Neurot	Height	-0.067	-0.114	-0.019	0.029	0.069	-0.023	0.157	0.14	0.055	202.57	-0.136	-0.234	-0.034	0.015
Neurot	IQ	-0.015	-0.062	0.031	0.619	0.105	0.008	0.199	0.03	0.092	783.88	-0.12	-0.225	-0.012	0.036
Neurot	Neurot	0.021	-0.033	0.074	0.506	0.109	0.018	0.202	0.025	0.1	-418.36	-0.088	-0.196	0.018	0.128
Neurot	SCZ	0.017	-0.034	0.067	0.571	-0.043	-0.141	0.058	0.399	-0.037	346.22	0.06	-0.051	0.172	0.309
Neurot	SRH	-0.063	-0.11	-0.017	0.044	0.005	-0.089	0.102	0.913	-0.002	108.43	-0.069	-0.178	0.039	0.236
SCZ	ADHD	0.036	-0.005	0.077	0.196	0.014	-0.06	0.089	0.706	0.019	61.13	0.022	-0.064	0.11	0.635
SCZ	BMI	0.062	0.021	0.102	0.026	0.031	-0.046	0.111	0.403	0.038	50.12	0.031	-0.058	0.118	0.501
SCZ	EA	-0.035	-0.08	0.01	0.218	0.011	-0.064	0.088	0.785	0	131.38	-0.046	-0.134	0.038	0.351
SCZ	Height	-0.006	-0.053	0.041	0.814	0.005	-0.059	0.071	0.887	0.003	181.97	-0.011	-0.092	0.068	0.799
SCZ	IQ	0.008	-0.033	0.049	0.757	0.032	-0.047	0.11	0.401	0.026	-276.71	-0.023	-0.111	0.066	0.616
SCZ	Neurot	0.049	0.008	0.089	0.074	-0.016	-0.089	0.058	0.669	-0.001	133.39	0.065	-0.014	0.146	0.164
SCZ	SCZ	-0.013	-0.054	0.026	0.622	-0.015	-0.091	0.058	0.685	-0.015	-13.6	0.002	-0.078	0.087	0.969
SCZ	SRH	-0.073	-0.113	-0.032	0.01	-0.003	-0.074	0.068	0.938	-0.019	96.1	-0.07	-0.155	0.014	0.128

Note. BMI = Body Mass Index; IQ = Intelligence; GCSE = General Certificate of Secondary Education (educational achievement); ADHD = Attention-Deficit/Hyperactivity Disorder; SCZ = Schizophrenia symptoms; EA = Educational Attainment; Neurot = Neuroticism; SRH = Self-rated Health; B = Between-family estimate; W = Within-family estimate; L.CI = Lower 95% bootstrap percentile interval; U.CI = Upper 95% bootstrap percentile interval; P = statistical significance of beta estimate; TotEff = Total effect derived as the intra-class correlation weighted sum of the within- and between family effect. PercRed = Reduction of prediction estimates when comparing within- to between-family estimates in percentage. BetaDiff = Difference of between-family and within-family beta coefficient; P.diff = statistical significance of difference between within- and between-family estimates.

Table S24. Within- and between-family prediction estimates based on polygenic scores with UK Biobank or all UK discovery samples removed

pheno	GPS	beta.B	L.CI.B	U.CI.B	P.B	beta.W	L.CI.W	U.CI.W	P.W	TotEff	PercRed	BetaDiff	Diff.L.CI	Diff.U.CI	P.diff
ADHD	BMI excl UKB	0.039	0.011	0.068	3.64e-02	0.016	-0.03	0.061	5.04e-01	0.023	60.52	0.024	-0.029	0.076	4.26e-01
ADHD	EA excl all UK	-0.095	-0.123	-0.065	2.93e-07	-0.093	-0.141	-0.045	2.44e-04	-0.094	1.15	-0.001	-0.057	0.056	9.72e-01
ADHD	Height excl UKB	0.005	-0.024	0.034	7.73e-01	0.02	-0.026	0.067	4.14e-01	0.015	-274.51	-0.015	-0.07	0.041	6.33e-01
BMI	BMI excl UKB	0.268	0.23	0.306	2.71e-26	0.292	0.234	0.352	2.18e-21	0.285	-9.07	-0.024	-0.096	0.047	5.34e-01
BMI	EA excl all UK	-0.107	-0.145	-0.067	2.15e-05	-0.069	-0.133	-0.006	4.56e-02	-0.081	35.51	-0.038	-0.111	0.037	3.73e-01
BMI	Height excl UKB	-0.012	-0.049	0.025	6.18e-01	-0.058	-0.121	0.007	8.12e-02	-0.043	-370.21	0.046	-0.028	0.118	2.70e-01
GCSE	BMI excl UKB	-0.085	-0.117	-0.053	7.53e-05	0.003	-0.036	0.04	8.87e-01	-0.048	103.15	-0.088	-0.136	-0.039	2.08e-03
GCSE	EA excl all UK	0.342	0.314	0.37	3.23e-64	0.137	0.098	0.177	2.15e-11	0.249	59.88	0.205	0.155	0.254	4.46e-13
GCSE	Height excl UKB	0.004	-0.027	0.034	8.58e-01	-0.01	-0.051	0.031	6.16e-01	-0.002	369.05	0.014	-0.037	0.064	6.33e-01
SRH	BMI excl UKB	-0.056	-0.09	-0.023	1.27e-02	0.004	-0.064	0.068	9.06e-01	-0.004	106.98	-0.06	-0.132	0.016	1.34e-01
SRH	EA excl all UK	0.051	0.016	0.085	2.51e-02	0.027	-0.047	0.1	4.66e-01	0.03	46.46	0.023	-0.057	0.105	5.88e-01
SRH	Height excl UKB	0.058	0.025	0.093	7.49e-03	0.007	-0.063	0.076	8.48e-01	0.014	88.27	0.052	-0.025	0.129	2.18e-01
Height	BMI excl UKB	-0.033	-0.071	0.007	2.04e-01	0.009	-0.042	0.062	7.39e-01	-0.009	127.78	-0.042	-0.108	0.024	2.65e-01
Height	EA excl all UK	0.032	-0.009	0.071	2.16e-01	0.009	-0.053	0.07	7.77e-01	0.019	73.04	0.023	-0.05	0.095	5.59e-01
Height	Height excl UKB	0.368	0.334	0.402	6.29e-54	0.35	0.296	0.404	4.03e-35	0.358	5.02	0.019	-0.047	0.083	6.05e-01
IQ	BMI excl UKB	-0.034	-0.072	0.002	1.65e-01	0.025	-0.028	0.079	3.44e-01	0	173.67	-0.06	-0.125	0.007	1.01e-01
IQ	EA excl all UK	0.211	0.175	0.249	1.27e-18	0.017	-0.042	0.075	5.63e-01	0.095	91.83	0.194	0.126	0.265	3.41e-07
IQ	Height excl UKB	-0.037	-0.074	0.001	1.33e-01	-0.007	-0.065	0.054	8.21e-01	-0.019	82.19	-0.03	-0.1	0.038	4.25e-01
Neurot	BMI excl UKB	0.003	-0.047	0.053	9.15e-01	-0.024	-0.118	0.07	6.11e-01	-0.021	825.01	0.027	-0.08	0.134	6.27e-01
Neurot	EA excl all UK	0.004	-0.044	0.053	8.88e-01	0.005	-0.093	0.105	9.17e-01	0.005	-25.51	-0.001	-0.112	0.109	9.85e-01

Neurot	Height excl UKB	-0.047	-0.093	0.001	1.12e-01	-0.011	-0.11	0.086	8.27e-01	-0.015	76.63	-0.036	-0.143	0.073	5.36e-01
SCZ	BMI excl UKB	0.068	0.028	0.107	1.40e-02	-0.01	-0.083	0.064	7.88e-01	0.01	114.1	0.077	-0.01	0.163	8.55e-02
SCZ	EA excl all UK	-0.073	-0.114	-0.032	6.90e-03	0.014	-0.061	0.092	7.31e-01	-0.008	118.91	-0.087	-0.178	0.002	7.22e-02
SCZ	Height excl UKB	0.019	-0.02	0.058	4.57e-01	-0.005	-0.086	0.073	8.92e-01	0.001	126.82	0.025	-0.062	0.112	5.96e-01

Note. BMI = Body Mass Index; IQ = Intelligence; GCSE = General Certificate of Secondary Education (educational achievement); ADHD = Attention-Deficit/Hyperactivity Disorder; SCZ = Schizophrenia symptoms; EA = Educational Attainment; Neurot = Neuroticism; SRH = Self-rated Health; UKB = UK Biobank; B = Between-family estimate; W = Within-family estimate; L.CI = Lower 95% bootstrap percentile interval; U.CI = Upper 95% bootstrap percentile interval; P = statistical significance of beta estimate; TotEff = Total effect derived as the intra-class correlation weighted sum of the within- and between family effect. PercRed = Reduction of prediction estimates when comparing within- to between-family estimates in percentage. BetaDiff = Difference of between-family and within-family beta coefficient; P.diff = statistical significance of difference between within- and between-family estimates.

Table S25 Phenotypic and polygenic score mean differences by polygenic score difference quantiles

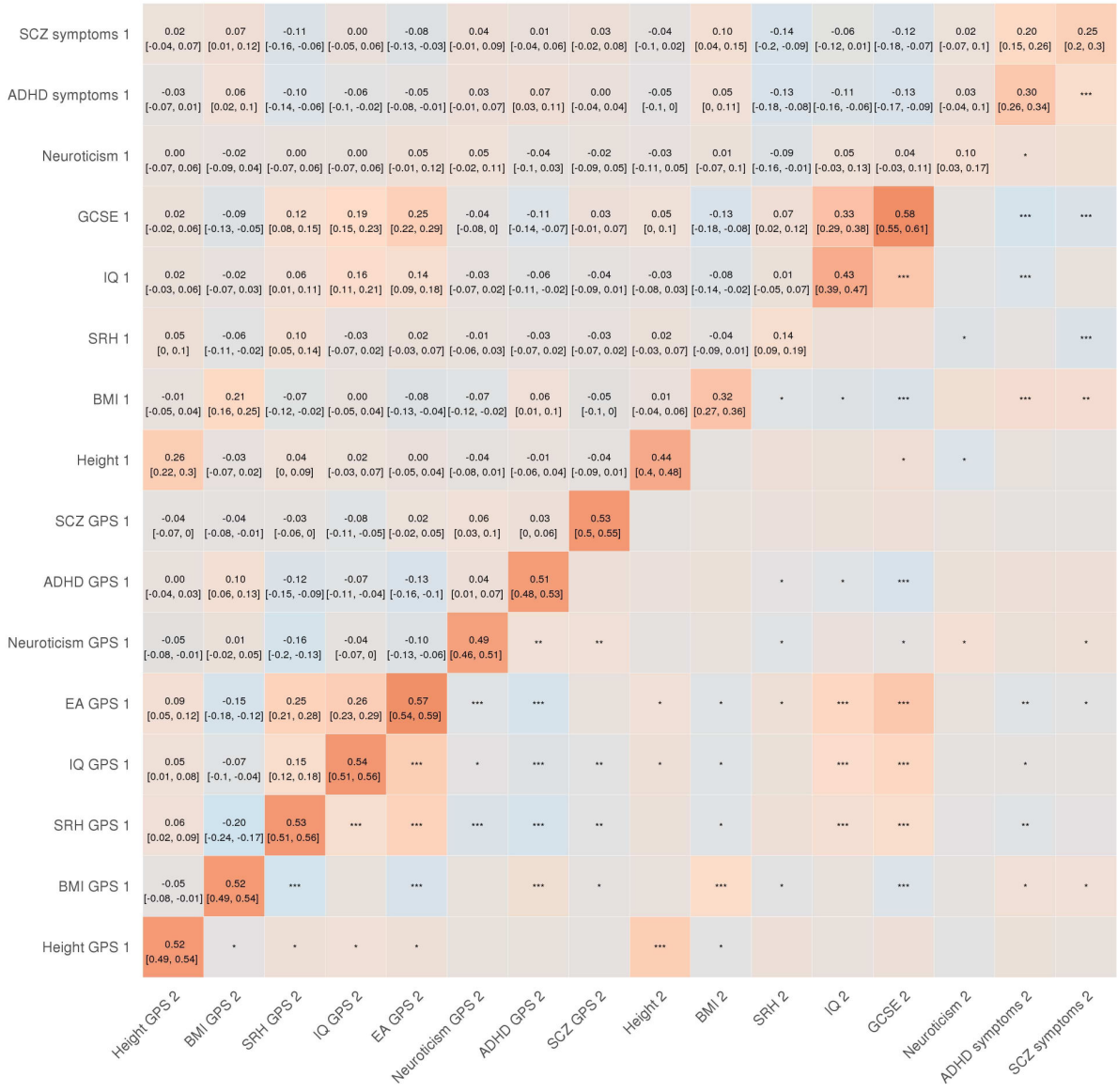
phenotype	GPS quant	mean phen	CI.L	CI.U	mean GPS	CI.L	CI.U
Height	1	0.270	-1.339	1.880	-0.004	-0.016	0.008
Height	2	1.718	0.203	3.233	-0.003	-0.033	0.027
Height	3	2.435	0.738	4.132	0.020	-0.030	0.071
Height	4	3.489	2.039	4.939	-0.030	-0.102	0.041
Height	5	1.433	-0.209	3.075	0.066	-0.025	0.158
Height	6	1.452	-0.144	3.048	0.019	-0.101	0.139
Height	7	4.620	2.862	6.379	-0.007	-0.154	0.141
Height	8	3.649	1.813	5.485	-0.038	-0.218	0.142
Height	9	7.021	5.316	8.727	0.065	-0.165	0.295
Height	10	8.989	7.181	10.798	-0.090	-0.422	0.242
BMI	1	0.064	-0.843	0.972	0.001	-0.012	0.014
BMI	2	0.968	0.126	1.810	0.004	-0.029	0.037
BMI	3	-0.555	-1.439	0.329	-0.011	-0.066	0.044
BMI	4	1.204	0.319	2.089	-0.027	-0.105	0.051
BMI	5	0.689	-0.073	1.451	0.055	-0.047	0.158
BMI	6	0.885	-0.009	1.778	-0.031	-0.159	0.097
BMI	7	1.495	0.584	2.406	0.075	-0.084	0.235
BMI	8	1.309	0.410	2.208	-0.084	-0.281	0.113
BMI	9	1.762	0.999	2.526	0.043	-0.200	0.286
BMI	10	2.933	2.092	3.773	-0.026	-0.378	0.327
IQ	1	1.369	-1.040	3.778	0.005	-0.006	0.016
IQ	2	-0.260	-2.860	2.341	-0.011	-0.042	0.019
IQ	3	1.139	-1.385	3.662	0.010	-0.041	0.061
IQ	4	-0.725	-3.287	1.838	-0.009	-0.081	0.063
IQ	5	2.081	-0.367	4.529	0.023	-0.069	0.115
IQ	6	1.483	-0.982	3.949	0.019	-0.099	0.137
IQ	7	1.288	-1.222	3.799	-0.051	-0.193	0.092
IQ	8	2.379	-0.215	4.973	0.092	-0.083	0.268
IQ	9	4.304	1.881	6.728	-0.101	-0.324	0.121
IQ	10	3.291	0.934	5.647	0.023	-0.293	0.340
GCSE	1	0.068	-0.067	0.203	0.005	-0.004	0.013
GCSE	2	-0.057	-0.204	0.090	0.001	-0.023	0.024
GCSE	3	0.195	0.053	0.336	-0.012	-0.051	0.026
GCSE	4	0.080	-0.063	0.223	0.021	-0.033	0.074
GCSE	5	0.097	-0.039	0.234	-0.022	-0.094	0.049
GCSE	6	0.266	0.129	0.403	0.054	-0.039	0.146
GCSE	7	0.226	0.083	0.369	0.084	-0.030	0.199
GCSE	8	0.307	0.180	0.435	0.060	-0.079	0.200
GCSE	9	0.341	0.192	0.490	0.013	-0.159	0.185

GCSE	10	0.471	0.327	0.614	-0.202	-0.445	0.041

Note. BMI = Body Mass Index; IQ = Intelligence; GCSE = General Certificate of Secondary Education (educational achievement); GPS = genome wide polygenic score; quant = quantile; phen = phenotype; CI.L = 95% lower confidence interval; CI.U = 95% upper confidence interval; GPS quant 1 = lowest absolute GPS twin pair difference quantile; GPS quant 10 = highest absolute GPS twin pair difference quantile.

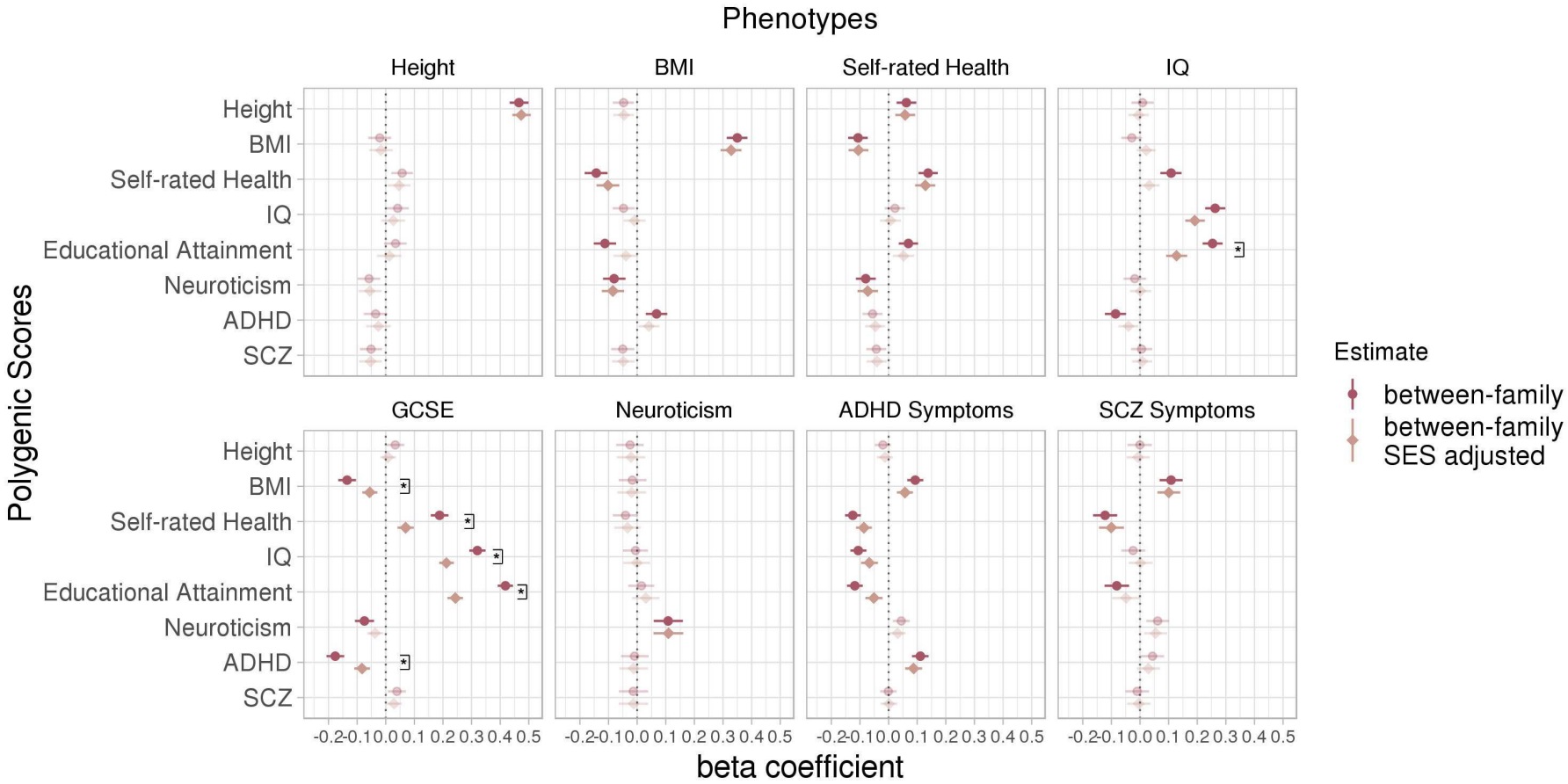
Supplementary Figures

Figure S1. Within-twin pair Pearson's correlation coefficients.



Note. BMI = Body Mass Index; IQ = Intelligence; GCSE = General Certificate of Secondary Education (educational achievement); ADHD = Attention-Deficit/Hyperactivity Disorder; SCZ = Schizophrenia; EA = Educational Attainment; SRH = Self-rated Health; 1 = Twin 1; 2 = Twin 2.

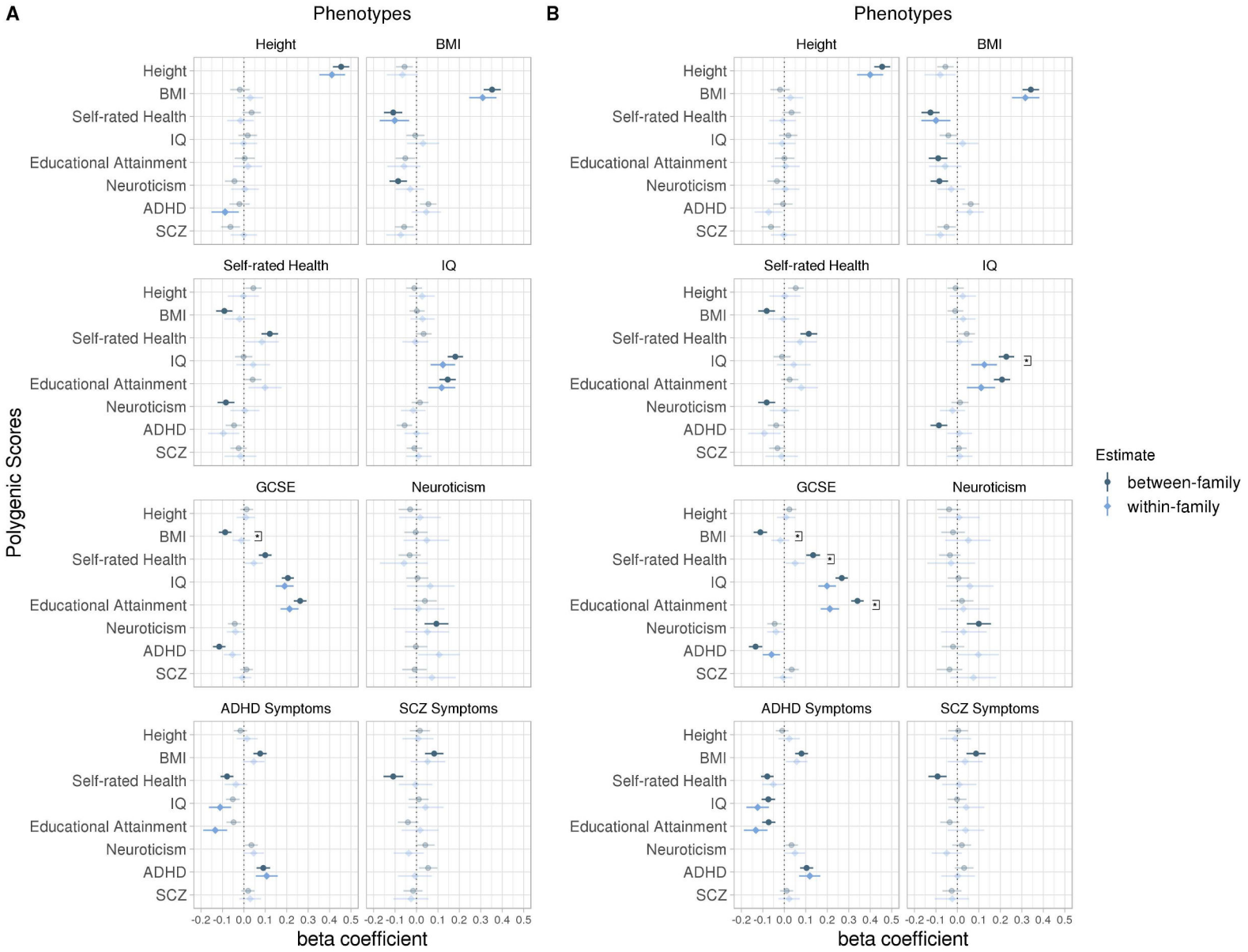
Figure S2. Effect sizes and significance of differences comparing between-family coefficients before and after controlling for family socio-economic status



Note. Between-family prediction estimates of eight outcomes using eight polygenic scores before and after statistical correction for family socio-economic status. Genome-wide Polygenic Scores are presented on the y-axis, predicting each of the eight phenotypic traits. Error bars

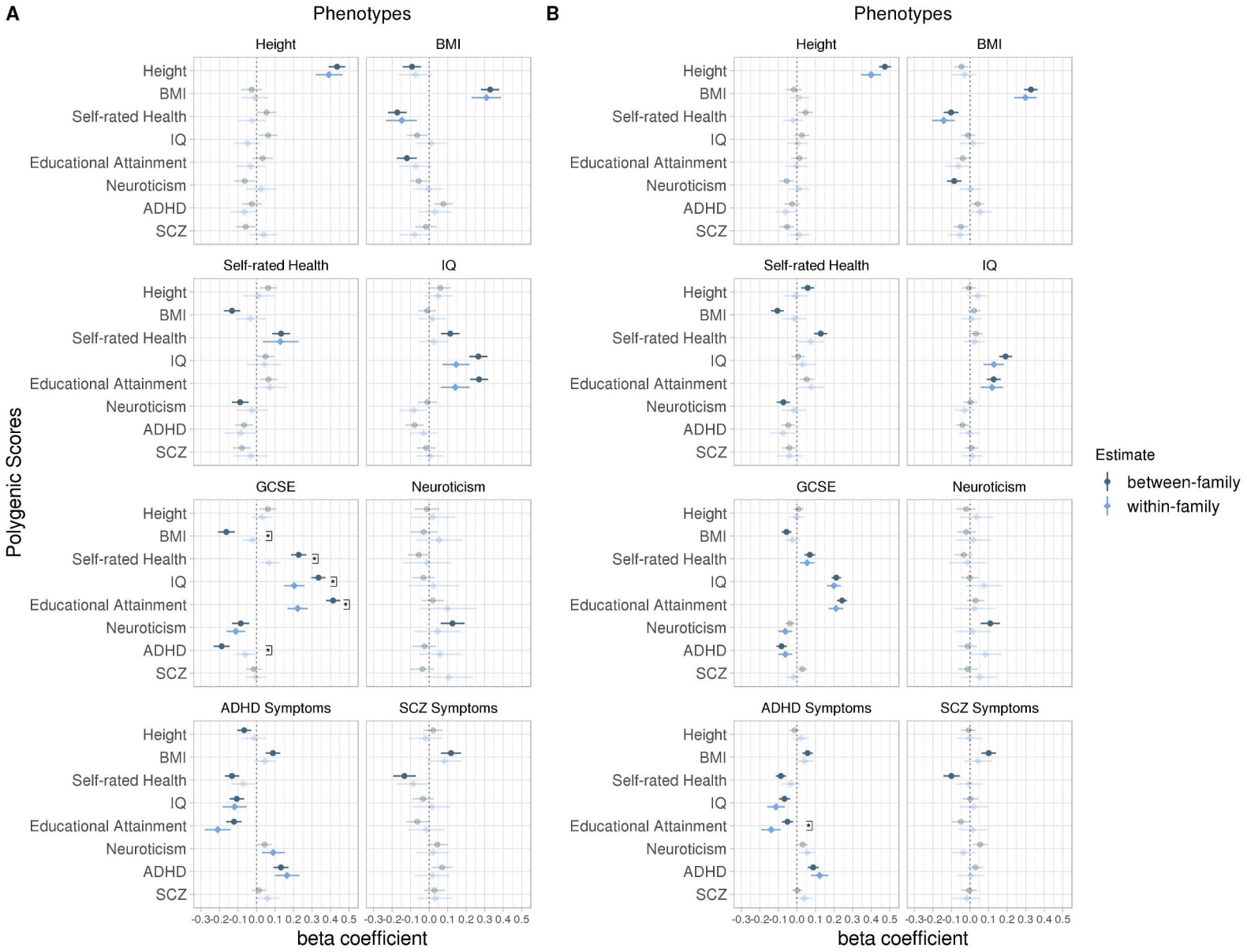
are 95% bootstrap percentile intervals based on 10,000 bootstrap samples (random resampling of DZ twin pairs with replacement). Opaque estimates indicate statistical significance at the false discovery rate corrected threshold of $p < 0.01$. Brackets indicate a significant difference between the two between-family prediction estimates, and only significant differences are shown where at least one of the estimates is significant at the false discovery rate corrected threshold of $p < 0.01$. The dotted line represents a beta coefficient of zero. BMI = Body Mass Index; IQ = Intelligence; GCSE = General Certificate of Secondary Education (educational achievement); ADHD = Attention-Deficit/Hyperactivity Disorder; SCZ = Schizophrenia.

Figure S3. Within- and between-family prediction estimates accounting for separate measures of family socio-economic status


A
B
Phenotypes
Height
BMI
Self-rated Health
IQ
GCSE
Neuroticism
ADHD Symptoms
SCZ Symptoms
Polygenic Scores
Educational Attainment
ADHD
SCZ
-0.2 -0.1 0.0 0.1 0.2 0.3 0.4 0.5
beta coefficient
Estimate
between-family
within-family

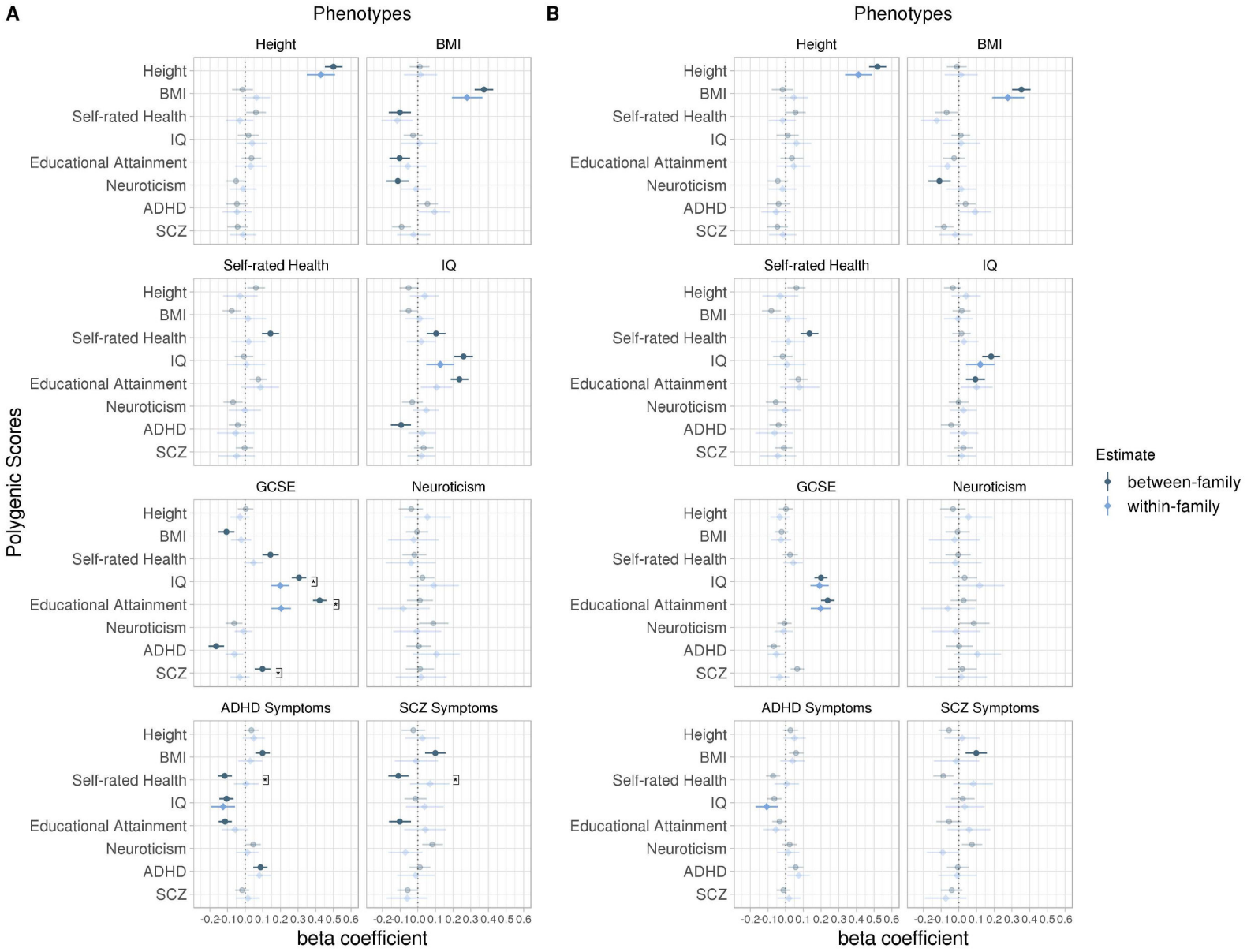
Note. Within- and between-family prediction estimates of eight outcomes using eight polygenic scores after statistically correcting for (A) parental education and (B) parental occupation. Genome-wide Polygenic Scores are presented on the y-axis, predicting each of the eight phenotypic traits. Error bars are 95% bootstrap percentile intervals based on 10,000 bootstrap samples (random resampling of DZ twin pairs with replacement). Opaque estimates indicate statistical significance at the false discovery rate corrected threshold of $p < 0.01$. Brackets indicate a significant difference between within- and between-family prediction estimate, and only significant differences are shown where at least one of the estimates is significant at the false discovery rate corrected threshold of $p < 0.01$. The dotted line represents a beta coefficient of zero. BMI = Body Mass Index; IQ = Intelligence; GCSE = General Certificate of Secondary Education (educational achievement); ADHD = Attention-Deficit/Hyperactivity Disorder; SCZ = Schizophrenia

Figure S4. Within- and between-family prediction estimates for same-sex twin pairs


A
Phenotypes
Height
BMI
Self-rated Health
IQ
GCSE
Neuroticism
ADHD Symptoms
SCZ Symptoms
Polygenic Scores
Height
BMI
Self-rated Health
IQ
Educational Attainment
Neuroticism
ADHD
SCZ
-0.3 -0.2 -0.1 0.0 0.1 0.2 0.3 0.4 0.5
beta coefficient
B
Phenotypes
Estimate
between-family
within-family

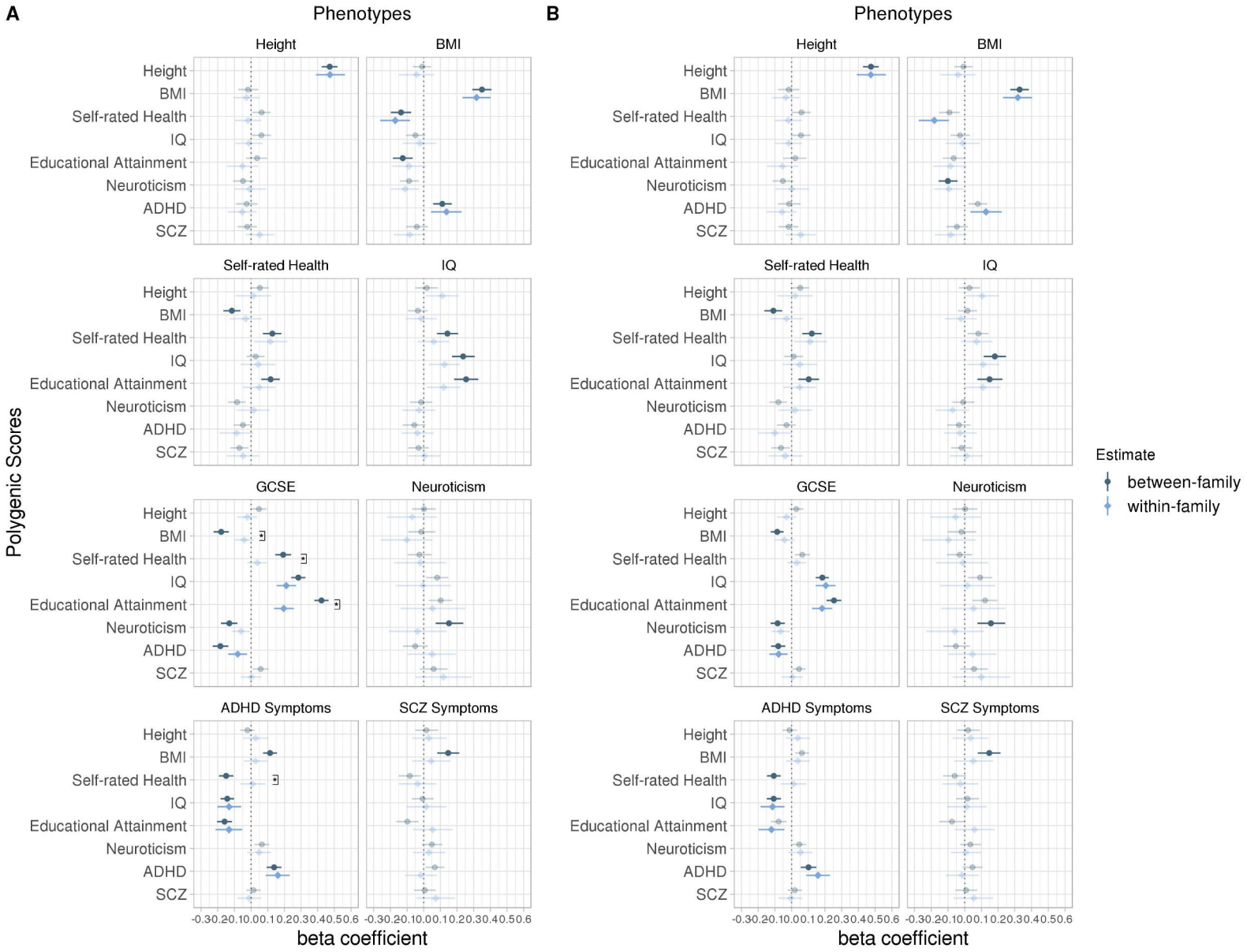
Note. Within- and between-family prediction estimates of eight outcomes using eight polygenic scores before (A) and after (B) statistical correction for family socio-economic status, based on same-sex twin pairs only. Genome-wide Polygenic Scores are presented on the y-axis, predicting each of the eight phenotypic traits. Error bars are 95% bootstrap percentile intervals based on 10,000 bootstrap samples (random resampling of DZ twin pairs with replacement). Opaque estimates indicate statistical significance at the false discovery rate corrected threshold of $p < 0.01$. Brackets indicate a significant difference between within- and between-family prediction estimate, and only significant differences are shown where at least one of the estimates is significant at the false discovery rate corrected threshold of $p < 0.01$. The dotted line represents a beta coefficient of zero. BMI = Body Mass Index; IQ = Intelligence; GCSE = General Certificate of Secondary Education (educational achievement); ADHD = Attention-Deficit/Hyperactivity Disorder; SCZ = Schizophrenia.

Figure S5. Within- and between-family prediction estimates for opposite-sex twin pairs


A
Phenotypes
B
Phenotypes
Height
BMI
Self-rated Health
IQ
GCSE
Neuroticism
ADHD Symptoms
SCZ Symptoms
Polygenic Scores
Educational Attainment
ADHD
SCZ
-0.2 -0.1 0.0 0.1 0.2 0.3 0.4 0.5 0.6
beta coefficient
Estimate
between-family
within-family

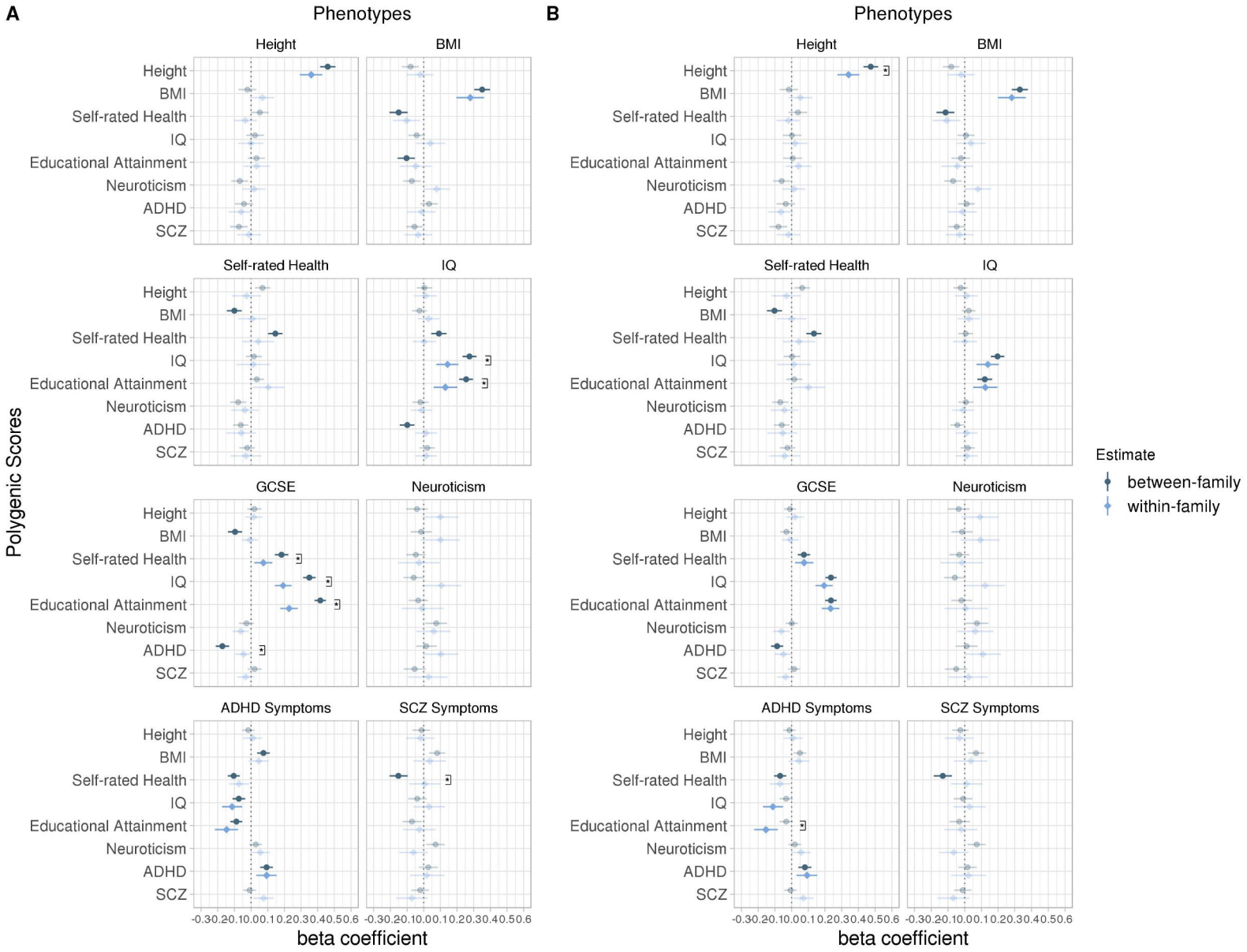
Note. Within- and between-family prediction estimates of eight outcomes using eight polygenic scores before (A) and after (B) statistical correction for family socio-economic status, based on opposite-sex twin pairs only. Genome-wide Polygenic Scores are presented on the y-axis, predicting each of the eight phenotypic traits. Error bars are 95% bootstrap percentile intervals based on 10,000 bootstrap samples (random resampling of DZ twin pairs with replacement). Opaque estimates indicate statistical significance at the false discovery rate corrected threshold of $p < 0.01$. Brackets indicate a significant difference between within- and between-family prediction estimate, and only significant differences are shown where at least one of the estimates is significant at the false discovery rate corrected threshold of $p < 0.01$. The dotted line represents a beta coefficient of zero. BMI = Body Mass Index; IQ = Intelligence; GCSE = General Certificate of Secondary Education (educational achievement); ADHD = Attention-Deficit/Hyperactivity Disorder; SCZ = Schizophrenia.

Figure S6. Within- and between-family prediction estimates based on twin pairs genotyped on OEE chip


A
B
Phenotypes
Height
BMI
Self-rated Health
IQ
GCSE
Neuroticism
ADHD Symptoms
SCZ Symptoms
Polygenic Scores
Educational Attainment
ADHD
SCZ
-0.3 -0.2 0.1 0.0 0.1 0.2 0.3 0.4 0.5 0.6
beta coefficient
Estimate
between-family
within-family

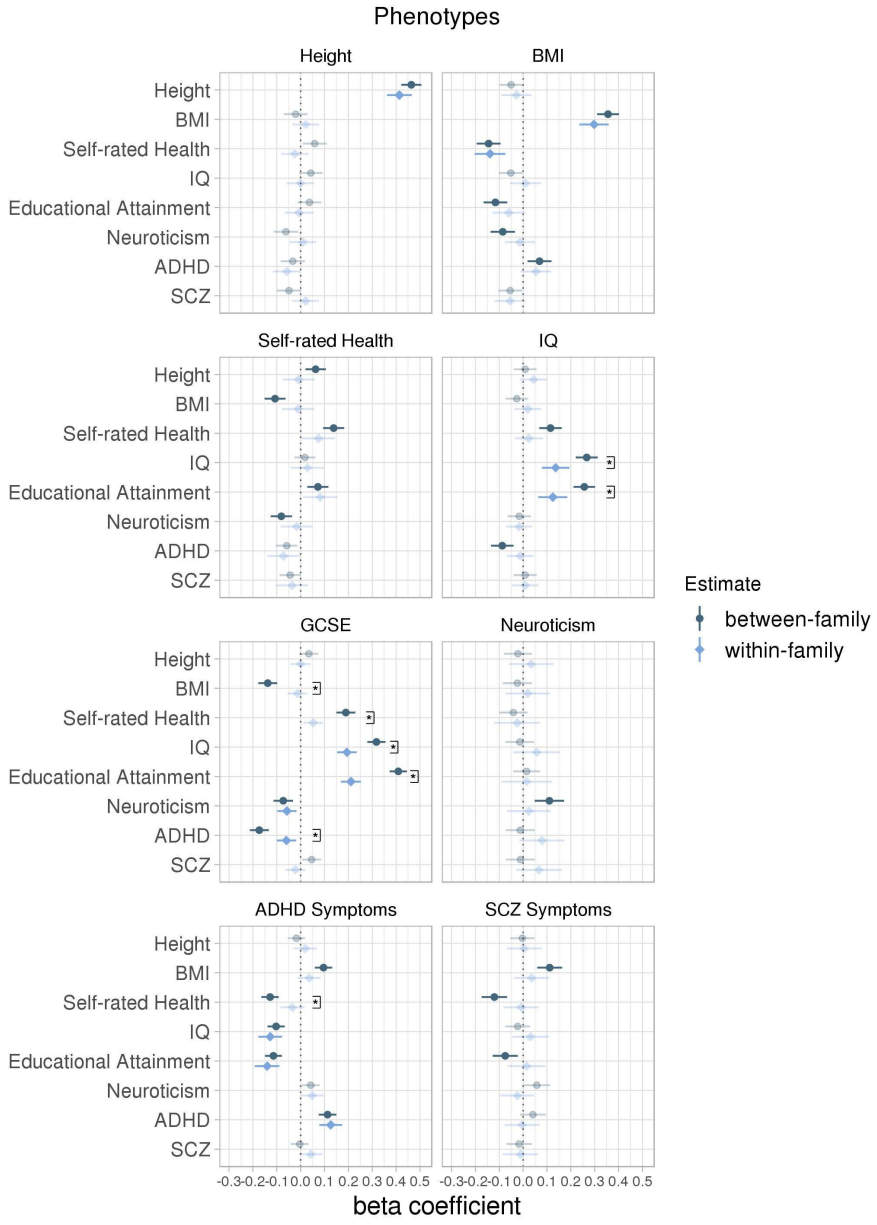
Note. Within- and between-family prediction estimates of eight outcomes using eight polygenic scores before (A) and after (B) statistical correction for family socio-economic status, based on twin pairs genotyped on OEE chip. Genome-wide Polygenic Scores are presented on the y-axis, predicting each of the eight phenotypic traits. Error bars are 95% bootstrap percentile intervals based on 10,000 bootstrap samples (random resampling of DZ twin pairs with replacement). Opaque estimates indicate statistical significance at the false discovery rate corrected threshold of $p < 0.01$. Brackets indicate a significant difference between within- and between-family prediction estimate, and only significant differences are shown where at least one of the estimates is significant at the false discovery rate corrected threshold of $p < 0.01$. The dotted line represents a beta coefficient of zero. BMI = Body Mass Index; IQ = Intelligence; GCSE = General Certificate of Secondary Education (educational achievement); ADHD = Attention-Deficit/Hyperactivity Disorder; SCZ = Schizophrenia.

Figure S7. Within- and between-family prediction estimates based on twin pairs genotyped on OEE chip and the co-twin genotyped on the Affymetrix chip


A
B
Phenotypes
Height
BMI
Self-rated Health
IQ
GCSE
Neuroticism
ADHD Symptoms
SCZ Symptoms
Polygenic Scores
Educational Attainment
ADHD
SCZ
-0.3 -0.2 -0.1 0.0 0.1 0.2 0.3 0.4 0.5 0.6
beta coefficient
Estimate
between-family
within-family

Note. Within- and between-family prediction estimates of eight outcomes using eight polygenic scores before (A) and after (B) statistical correction for family socio-economic status, based on twin pairs with one twin genotyped on OEE chip and the other twin genotyped on Affymetrix chip. Genome-wide Polygenic Scores are presented on the y-axis, predicting each of the eight phenotypic traits. Error bars are 95% bootstrap percentile intervals based on 10,000 bootstrap samples (random resampling of DZ twin pairs with replacement). Opaque estimates indicate statistical significance at the false discovery rate corrected threshold of $p < 0.01$. Brackets indicate a significant difference between within- and between-family prediction estimate, and only significant differences are shown where at least one of the estimates is significant at the false discovery rate corrected threshold of $p < 0.01$. The dotted line represents a beta coefficient of zero. BMI = Body Mass Index; IQ = Intelligence; GCSE = General Certificate of Secondary Education (educational achievement); ADHD = Attention-Deficit/Hyperactivity Disorder; SCZ = Schizophrenia

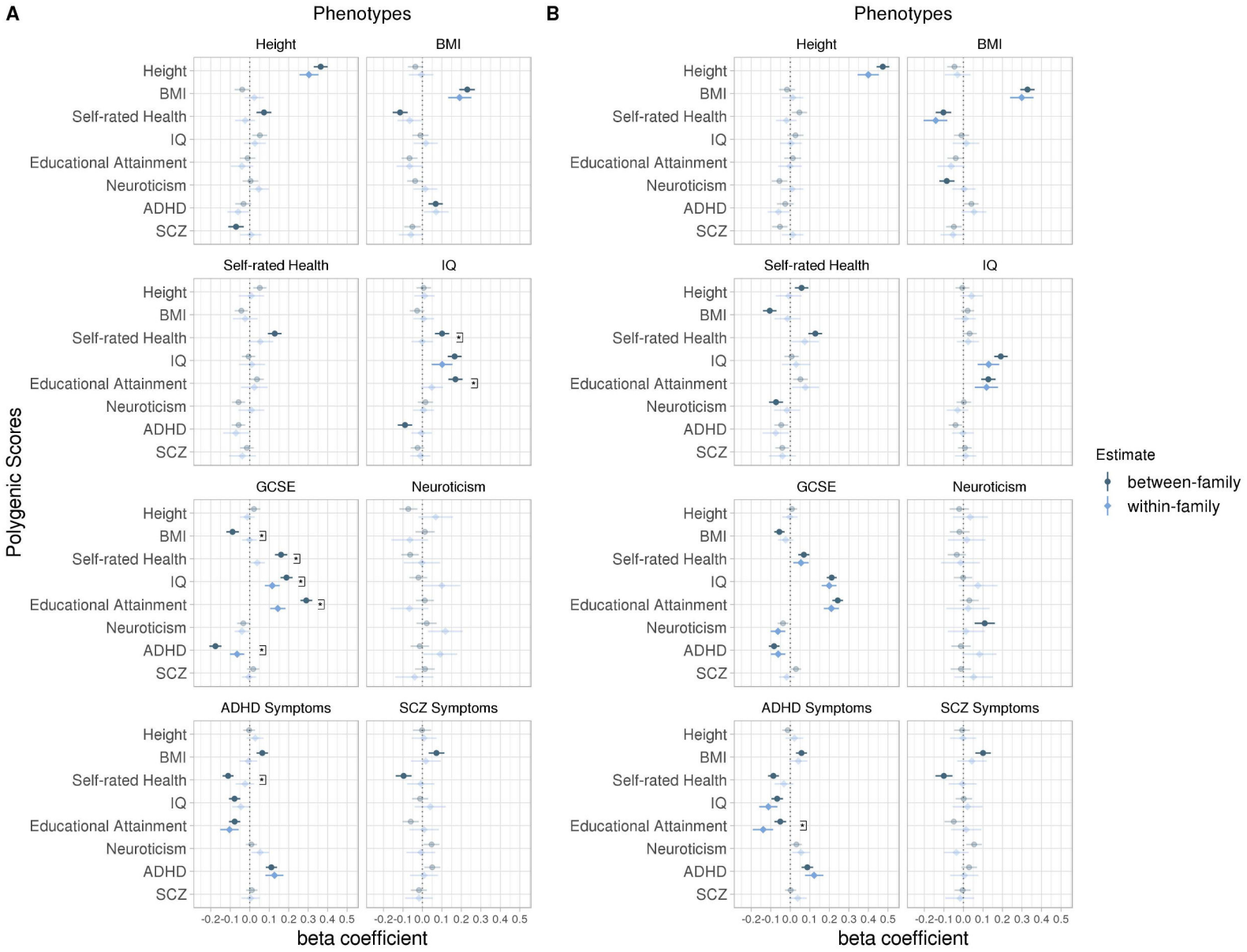
Figure S8. Within- and between-family coefficients estimated using a SNP-kinship matrix as random effect



Note. Within- and between-family prediction estimates of eight outcomes using eight polygenic scores. Genome-wide Polygenic Scores are presented on the y-axis, predicting each of the eight phenotypic traits. Error bars are 95% confidence intervals calculated using the standard errors of the fixed effects. Opaque estimates indicate statistical significance at the false discovery rate corrected threshold of $p < 0.01$. Brackets indicate a significant difference between within- and between-family prediction estimate, and only significant differences are shown where at least one of the estimates is significant at the false discovery rate corrected threshold of $p < 0.01$. The dotted line represents a beta coefficient of zero. BMI = Body Mass Index; IQ = Intelligence; GCSE = General Certificate of Secondary Education

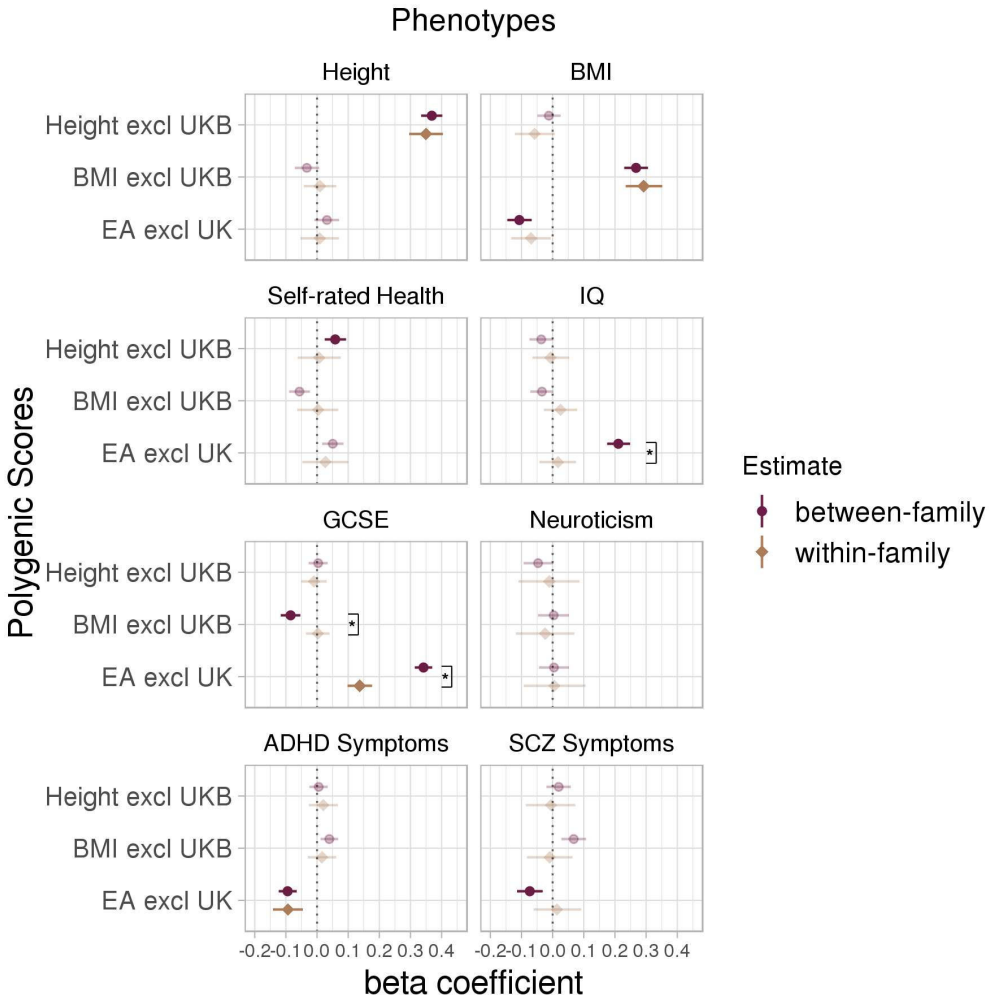
(educational achievement); ADHD = Attention-Deficit/Hyperactivity Disorder; SCZ = Schizophrenia.

Figure S9. Within- and between-family prediction estimates using polygenic scores calculated with causal fraction of markers of 0.1


A
B
Phenotypes
Height
BMI
Self-rated Health
IQ
GCSE
Neuroticism
ADHD Symptoms
SCZ Symptoms
Polygenic Scores
Educational Attainment
ADHD
SCZ
-0.2 -0.1 0.0 0.1 0.2 0.3 0.4 0.5
beta coefficient
Estimate
between-family
within-family

Note. Within- and between-family prediction estimates of eight outcomes using eight polygenic scores before (A) and after (B) statistical correction for family socio-economic status, using polygenic scores calculated with causal fraction of markers of 0.1. Genome-wide Polygenic Scores are presented on the y-axis, predicting each of the eight phenotypic traits. Error bars are 95% bootstrap percentile intervals based on 10,000 bootstrap samples (random resampling of DZ twin pairs with replacement). Opaque estimates indicate statistical significance at the false discovery rate corrected threshold of $p < 0.01$. Brackets indicate a significant difference between within- and between-family prediction estimate, and only significant differences are shown where at least one of the estimates is significant at the false discovery rate corrected threshold of $p < 0.01$. The dotted line represents a beta coefficient of zero. BMI = Body Mass Index; IQ = Intelligence; GCSE = General Certificate of Secondary Education (educational achievement); ADHD = Attention-Deficit/Hyperactivity Disorder; SCZ = Schizophrenia.

Figure S10. Within- and between-family prediction estimates based on polygenic scores with UK Biobank or all UK discovery samples removed



Note. Within- and between-family prediction estimates of eight outcomes in using polygenic scores with UK Biobank or all UK samples removed at discovery stage. Genome-wide Polygenic Scores are presented on the y-axis, predicting each of the eight phenotypic traits. Error bars are 95% bootstrap percentile intervals based on 10,000 bootstrap samples (random resampling of DZ twin pairs with replacement). Opaque estimates indicate statistical significance at the false discovery rate corrected threshold of $p < 0.01$. Brackets indicate a significant difference between within- and between-family prediction estimate, and only significant differences are shown where at least one of the estimates is significant at the false discovery rate corrected threshold of $p < 0.01$. The dotted line represents a beta coefficient of zero. BMI = Body Mass Index; IQ = Intelligence; GCSE = General Certificate of Secondary Education (educational achievement); ADHD = Attention-Deficit/Hyperactivity Disorder; SCZ = Schizophrenia; EA = Educational Attainment; UKB = UK Biobank.

References

1. Vilhjalmsson, B.J., Yang, J., Finucane, H.K., Gusev, A., Lindstrom, S., Ripke, S., Genovese, G., Loh, P.-R., Bhatia, G., Do, R., et al. (2015). Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. Am. J. Hum. Genet. 97, 576–592.

2. Yengo, L., Sidorenko, J., Kemper, K.E., Zheng, Z., Wood, A.R., Weedon, M.N., Frayling, T.M., Hirschhorn, J., Yang, J., Visscher, P.M., et al. (2018). Meta-analysis of genome-wide association studies for height and body mass index in similar to 700 000 individuals of European ancestry. Hum. Mol. Genet. 27, 3641–3649.

3. McInnes, G., Tanigawa, Y., DeBoever, C., Lavertu, A., Olivieri, J.E., Aguirre, M., and Rivas, M.A. (2018). Global Biobank Engine: enabling genotype-phenotype browsing for biobank summary statistics. Bioinformatics 9, 1612.

4. Harris, S.E., Hagenaars, S.P., Davies, G., Hill, W.D., Liewald, D.C.M., Ritchie, S.J., Marioni, R.E., Sudlow, C.L.M., Wardlaw, J.M., McIntosh, A.M., et al. (2017). Molecular genetic contributions to self-rated health. Int J Epidemiol 46, 994–1009.

5. Savage, J.E., Jansen, P.R., Stringer, S., Watanabe, K., Bryois, J., de Leeuw, C.A., Nagel, M., Awasthi, S., Barr, P.B., Coleman, J.R.I., et al. (2018). Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. Nat. Genet. 50, 912–919.

6. Allegrini, A., Selzam, S., Rimfeld, K., Stumm, von, S., Pingault, J.-B., and Plomin, R. (2019). Genomic prediction of cognitive traits in childhood and adolescence. Molecular Psychiatry.

7. Lee, J.J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., Nguyen-Viet, T.A., Bowers, P., Sidorenko, J., Linner, R.K., et al. (2018). Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. Nat. Genet. 50, 1112–1121.

8. Luciano, M., Hagenaars, S.P., Davies, G., Hill, W.D., Clarke, T.-K., Shirali, M., Harris, S.E., Marioni, R.E., Liewald, D.C., Fawns-Ritchie, C., et al. (2018). Association analysis in over 329,000 individuals identifies 116 independent variants influencing neuroticism. Nat. Genet. 50, 6–11.

9. Demontis, D., Walters, R.K., Martin, J., Mattheisen, M., Als, T.D., Agerbo, E., Baldursson, G., Belliveau, R., Bybjerg-Grauholm, J., Bækvad-Hansen, M., et al. (2019). Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. Nat. Genet. 51, 63–75.

10. Pardiñas, A.F., Holmans, P., Pocklington, A.J., Escott-Price, V., Ripke, S., Carrera, N., Legge, S.E., Bishop, S., Cameron, D., Hamshere, M.L., et al. (2018). Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. Nat. Genet. 50, 381–389.